

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

FACULTY OF HUMANITIES

Philosophy

Inference and Action: Relating beliefs to the World

by

Javier González de Prado Salas

Thesis for the degree of Doctor of Philosophy

July 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF HUMANITIES

Philosophy

Thesis for the degree of Doctor of Philosophy

INFERENCE AND ACTION: RELATING BELIEFS TO THE WORLD

Javier González de Prado Salas

The goal of this dissertation is to offer a practice-based account of intentionality. My aim is to examine what sort of practices agents have to engage in so as to count as talking and thinking *about* the way the world is – that is, what sort of practices count as representational.

Representational practices answer to the way the world is: what is correct within such practices depends on the way things are, rather than (only) on the attitudes of agents. An account of representation must explain how such objective standards of correctness are introduced in human practices: one must explain how the world gets to have a say in what is correct in human discursive practices.

Roughly, my proposal is that human discursive practices become responsive to the way things are by virtue of involving practical interactions with the world. The outcomes of these interactions depend on the way the world is and the evaluation of such outcomes contributes to determining which moves within the practice count as correct. Due to our practical engagement with the environment, thus, the world gets to constrain discursive practices.

In order to flesh out my proposal, I develop a practice-based characterization of intentional or representational content. On this sort of approach, expressing intentional contents is seen as a matter of playing a certain role in relevant practices, rather than as a matter of engaging in some word-world relation. The expression of content, thus, is

explained in terms of use. In particular, I adopt an inferentialist perspective, according to which discursive moves express contents because of their role in practices of giving and asking for reasons. I investigate how practical engagement with the environment introduces friction with the world in these practices of giving and asking for reasons.

One of the main conclusions reached in the dissertation is that defeasibility is an essential feature of objective representational practices – so that attributions of representational correctness are revisable in an open-ended way. The discussion of defeasible reasoning – and of the way in which defeasibility shapes human representational practices – is a central point of this dissertation.

Table of Contents

Table of Contents	i
DECLARATION OF AUTHORSHIP	v
Acknowledgements	vii
Chapter 1: Introduction	1
Chapter 2: A Practice-Based Approach to Intentionality	7
2.1 Introduction	7
2.2 The problem of representation	7
2.3 A pragmatist view of intentionality	10
2.3.1 Representationalism and pragmatism	10
2.3.2 Practice-based semantics	16
2.4 Objectivity as answering to the world	20
2.5 Pragmatism, local and global	27
2.6 Conclusions	32
Chapter 3: Inferentialism as a Practice-Based Semantics	35
3.1 Introduction	35
3.2 A pragmatist motivation for inferentialism	36
3.3 Inferentialism as a practice-based semantics	39
3.3.1 Inference and practices of giving and asking for reasons	39
3.3.2 Commitments, entitlements and the expression of content	41
3.3.3 Commitment to a claim as a commitment to doing something	45
3.3.4 Some features of the practices of giving and asking for reasons	46
3.3.5 Summary	47
3.4 Pragmatism and the normative dimension of discursive practices: normative statuses and normative attitudes	48
3.5 The semantics of inferentialism	51
3.5.1 Material validity	51
3.5.2 Inferential non-monotonicity	54

3.5.3	Commitment-preserving and entitlement-preserving inferences	55
3.6	Inferentialism and semantic holism	59
3.7	Inferentialism and objectivity	61
3.8	Conclusions	67
Chapter 4:	Perception and Action in Discursive Practices	69
4.1	Introduction	69
4.2	Perceptual entitlement.....	70
4.3	Intentional agency and practical reasoning	75
4.3.1	Intentional agency	75
4.3.2	Practical reasons, facts and attitudes.....	76
4.3.3	The expressive role of attributions of pro-attitudes and goals.....	79
4.3.4	Objections to the expressive role of pro-attitudes	82
4.3.5	Practical inference	84
4.4	Intentional agency and success	89
4.5	Brandom’s TOTE-cycles.....	92
4.6	Practical success and the updating of representational views.....	95
4.7	Conclusions	97
Chapter 5:	Success Semantics	99
5.1	Introduction	99
5.2	Success semantics	100
5.3	Is truth necessary for success?	102
5.4	Is truth sufficient for success?.....	104
5.4.1	False collateral beliefs.....	104
5.4.2	Ignored obstacles.....	108
5.5	Truth as a necessary condition for success: revisiting the question	117
5.6	Truth without guarantee of success	118
5.6.1	Success despite falseness	119
5.6.2	Ignored obstacles again	122
5.7	Conditions of success.....	124
5.7.1	Reductionism	125

5.7.2	Holism	126
5.8	Conclusions	127
Chapter 6:	Biological Goals	131
6.1	Introduction	131
6.2	Practical success as the fulfilment of biological goals	132
6.3	Proliferation and self-maintenance as evaluative standards.....	135
6.3.1	Evolutionary and organizational accounts of biological goals	135
6.3.2	A challenge for evolutionary and organizational accounts	137
6.4	Selection and evaluation	141
6.5	Self-maintenance as the source of teleology	145
6.6	Conclusions	148
Chapter 7:	The Defeasibility of Practical Reasoning Error! Bookmark not defined.	
7.1	Introduction	149
7.2	Ignored obstacles	151
7.3	Ignored obstacles and the problem of indeterminacy.....	155
7.4	Defeasible reasoning	159
7.5	Defeasibility and practical reasoning.....	164
7.6	Defeasibility as a source of inferential occasion-sensitivity.....	167
7.7	Defeasibility and representation.....	171
7.8	Conclusions	173
Chapter 8:	Representational Practices and Environmental Constraints.....	175
8.1	Introduction	175
8.2	Self-regulation and goal-directed behavior	177
8.3	Making room for error.....	181
8.3.1	Error and alternative perspectives	181
8.3.2	Objectivity and defeasibility	184
8.3.3	Objective representation and fallibility	187
8.4	Objectivity and relativism	191
8.5	The link with the world.....	195
8.6	Conclusion.....	201

Chapter 9: Conclusions	203
List of References	209

DECLARATION OF AUTHORSHIP

I, Javier González de Prado Salas

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Inference and Action: Relating Beliefs to the World

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed:Javier González de Prado Salas.....

Date:01-12-2015.....

Acknowledgements

I would like to thank Daniel Whiting for his willingness to supervise this project and for his constant advice, support and rigorous, but always constructive and encouraging feedback. What I have learnt under his guidance goes far beyond what this dissertation might manage to show.

I also have to thank the Philosophy staff and students at the University of Southampton for creating a supportive and stimulating environment that has allowed me to develop my research in the best possible conditions.

This dissertation has been an international project, if only because it has been written in different countries. My research could not have taken place without the encouragement and support of the department of Logic, History and Philosophy of Science of Universidad Nacional de Educación a Distancia (UNED, Madrid). In particular, my work has greatly benefited from advice and discussions with David Teira, Jesús Zamora Bonilla and Cristian Saborido. Additional thanks are due to David Teira for his invaluable help with administrative issues at UNED. I must also thank philosophers at other departments in Madrid (especially Jesús Vega, Ignacio Vicario and Mauricio Suárez).

A fundamental part of a PhD project is the graduate community in which it is developed. Thus, I have to thank all the graduate students that in different ways have helped me while I was working in this dissertation – particularly (but not only) in Madrid and Southampton. Among many others, I should mention María Muñoz, Álex Díaz, Manuel Heras Escribano, Isabel Gamero, Miranda del Corral, José Ángel Gascón, Marco Antonio Joven, Susana Monsó, Tsung-Hsing Ho and Ivan Milic.

Several sections of this dissertation have been presented at different conferences and workshops. I thank the audiences for their feedback. I also want to thank Cristian Saborido and Manuel Heras Escribano for their comments about the material in chapter 6. I have learnt a lot about normativity and related topics from discussions with Víctor Verdejo and Miranda del Corral.

Although it goes without saying, I am especially grateful to Natasha, my family and friends.

Chapter 1: Introduction

We talk and think about things in the world. The correctness or truth of our assertions and beliefs depends on the way the world is. For example, if I say that it is raining outside, the correctness of my assertion will depend on the meteorological state of the world – on whether it is actually raining outside.

Our speech and thought are in this way capable of describing the world. What makes it so? What makes our talking and thinking count as being directed towards the things around us? Although describing our surroundings – describing the weather outside or the contents of my fridge – is an ordinary aspect of our lives, it is nonetheless one of the most distinctive and intriguing human capacities. In the non-biological world, one does not seem to find this sort of representational relations – by virtue of which certain events or states count as correct or incorrect depending on the state of other parts of the world. The question I will address in this dissertation is how human speech and thought get to incorporate this representational dimension. In other words, I will investigate how our utterances and beliefs get to be *about* the world.

I will approach human language and thought as an activity – something humans do. So, my aim will be to discuss what sort of practices individuals have to be engaged in so as to count as talking and thinking about the way the world is – that is, what sort of practices count as representational. An essential aspect of representational practices, as I will understand them, is their objectivity: representational practices purport to answer to how the world they represent is. What is correct in such practices depends on the way things are in the world, rather than on the opinions of the participants in the practices. I will examine how this friction with the world arises in human practices of talking and thinking (or, as I will refer to them, discursive practices).

My project can be seen as related to the work of authors associated with what is known as new pragmatism (among others, Robert Brandom (1994, 2008), Huw Price (2011), Michael Williams (2001), Jeffrey Stout (2007), Rebecca Kukla and Mark Lance (2009, 2014) and John Haugeland (1998)). These authors also study language and thought as embedded in human practices, while trying to allow for the possibility of such practices being objective – that is, for the possibility of such practices responding to the way the

Chapter 1

world actually is. Such pragmatist approaches face the challenge of accounting for the existence of objective standards of correctness which are, nevertheless, established by the activity of the participants in the practice. How does the world acquire this sort of normative authority over human practices?

The thesis I want to explore, in a nutshell, is that certain practices can be said to be representational (in the sense that what is correct in the practice is determined by how the world represented is) because the practical outcomes of interactions with the environment are taken into account by agents in order to evaluate the correctness of moves within the practice. My proposal will be that our discursive practices are constrained by the way the world is by virtue of involving practical interactions with the environment. The outcomes of these interactions depend on how things stand in the world, and such outcomes have a say in determining the correctness of moves within the practice. When assessing what is correct within discursive practices, agents take into account the results of the practical dealings with the environment involved in those practices – and the way things are constrains, in turn, the results of these dealings. Thus, friction with the world is introduced in our discursive practices by virtue of our practical engagement with the environment.

According to this proposal, the world can be said to have authority over our practices, but it is an authority by deferral: by turning to the outcomes of our practical interactions with the environment in order to determine the correctness of our views, we bestow such authority on the world. Ultimately, standards of correctness are instituted by the individuals participating in the relevant practices, but such individuals may choose to defer to the world in order to determine what counts as correct in the practice. Once authority is delegated in such a way, it is no longer up to the participants in the practice to decide what is correct and what is incorrect. The resulting practices will, in this way, answer to how things are in the world, rather than to the attitudes of individuals. Thus, this sort of practices will allow for objective representation of the world. I will not claim that practical interactions with the world represented are necessary for objective representation, but only that one way in which objective representational practices may arise is by virtue of such interactions – and moreover, that this is how our mundane representational discourse about middle sized things around us actually arises. I will not have anything to say about whether other kind of representational practices, not relying

on practical engagement with the world, are possible. My proposal, therefore, aspires to offer a practice-based account of representation. In this sort of account, representational content is taken to be conferred by practices. Expressing some representational content – describing the world in some way – would be a matter of playing a certain role in a relevant practice. The most prominent alternatives to this practice-based approach are views in which representation is accounted for in terms of some word-world relation, some relation connecting our talk and thought with the world (such as reference or correspondence) – perhaps providing some reduction of such relations to causal ones. The pragmatist approach I will explore does not deny that our thought and talk stands in representational relations to the world. However, representation is not ultimately analysed in terms of such relations: instead, what it is to represent the world – what it is to stand in representational relations – is taken to be a matter of engaging in certain practices. So, one would focus on the place occupied by representational activities in our practices, and in particular on how agents use words – how they talk and think – when they purport to represent the world. I will take inferentialism as my framework for developing a practice-based semantics (more specifically, the version of inferentialism developed by Robert Brandom (1994, 2000)). According to inferentialism, the contents of claims are determined by the inferential entailments they are involved in. Inferentialism can be naturally regarded as a practice-based semantics because the inferential role of claims is straightforwardly related to the role that discursive moves expressing such claims play in reasoning and in argumentative practices (which Brandom (1994) calls practices of giving and asking for reasons, following Sellars (1951)). For instance, the claim “There are apples in the bowl” entails the claim “There is fruit in the bowl”; accordingly, having reasons to accept that there are apples in the bowl will give you reasons to accept that there is fruit in the bowl. If it is correct for you to assert the first claim, it will be correct for you to assert the second one. In chapter 3 will discuss in detail the main features of inferentialist semantics and its relations to language use.

Crucially for my project, practices of giving and asking for reasons include – at least when understood in a broad way – interactions with the environment, through action and perception (see Brandom 1994: ch. 4). So, one may acquire reasons to accept some claim by virtue of being in a certain perceptual state – by virtue of exercising reliable perceptual

Chapter 1

responsive sensibilities. For instance, seeing a red table in front of me, in good light conditions, entitles me to claiming that there is red table in front of me. Similarly, having reasons to accept certain claims may provide reasons to endorse (or reject) some course of action. In this way, the fact that some mushroom is poisonous is a good reason not to eat it, and the fact that it is raining is a good reason for opening the umbrella.

I will argue that when a good piece of practical reasoning recommends some action as a sufficient means to its end, the appropriate performance of the action will necessarily succeed in achieving its goal. Thus, if an action fails to bring about the achievement of its goal, then the piece of reasoning recommending it as a sufficient means for such a goal will turn out to be an actually bad piece of practical reasoning. Either the piece of reasoning is unsound (its premises are false, there is no entitlement to endorsing them) or it is invalid (the inferential link underlying it is not correct). In this way, by evaluating the success of their actions, agents may revise and update their views – the claims they accept and the inferential links they endorse. Since the success of actions, as I have pointed out, depends on the way things are, the world will constrain this process of revision and updating.

The connection between practical success and representational accuracy, however, is less direct than what may seem at first sight. On the one hand, an action derived only from true beliefs may fail because of the presence of ignored obstacles (for instance, even if you correctly believe that the bottle contains water, you may fail in drinking from it if the bottle cap is stuck). Typically, the number of possible obstacles will be indefinitely large, so it is not feasible to explicitly discard each of them. As a consequence, the inferences underpinning practical reasoning are defeasible (i.e. non-monotonic): they may become bad due to the presence of ignored obstacles, which would thwart the success of the actions recommended. On the other hand, agents may always be mistaken in assessing some action as successful or failed (for instance, an agent may assess such success perceptually in conditions where her perception is not reliable). Therefore, the fact that an action is assessed as failed does not entail that the agent had inaccurate beliefs: perhaps there was some unexpected obstacle, or the evaluation is mistaken. In this sense, the evaluation of the outcomes of actions is also defeasible.

Thus, in general representational views (views about how things are in the world) can only be endorsed in a revisable way: in principle, one should always be open to the

possibility of such views turning out to be mistaken. This does not mean, however, that objectivity is an ideal that can only be approximated, but never reached in actual human practices. On the contrary, defeasibility allows for the combination of the objective nature of representation with human fallibility. Although there is always the possibility that our views are mistaken due to some ignored fact, that does not imply that such views cannot be said to be correct – and answer to the way the world is – in cases when things actually go well and such defeaters are absent. Agents do not need to be able to rule out each possible defeater in order for their views to count as genuinely entitled. I will argue, in this way, that objective representation is possible even if our representational engagement with the world is precarious and fallible. The defeasibility of human reasoning and, generally, of human representational practices will be one of the main topics of this dissertation, and will be discussed at length in chapters 5, 7 and 8.

The outline of the dissertation is as follows. In chapter 2 I present in more detail the project undertaken in the dissertation, that is, the project of accounting for the representational dimension of human discursive practices. I discuss the sort of pragmatist or practice-based approach to intentionality I will endorse, and I examine the constraints that such an approach imposes on semantic theories. I describe the challenge faced by these pragmatist approaches in making room for objectivity, and I sketch how I plan to meet such a challenge.

In chapter 3 I introduce Brandom's inferentialist semantics as the practice-based semantic framework in which my project will be developed. I present an overview of the main tenets of inferentialism and I examine some problematic aspects of Brandom's proposal – focusing on those that are more relevant for the purposes of the dissertation. I also indicate how I will depart from Brandom in those problematic points.

In chapter 4, I discuss how interactions with the world – both through perception and through action – are integrated in an inferentialist model. I examine the role played by evaluations of success in discursive practices, and I explore the way in which such evaluations of success lead to revising and updating the views endorsed by agents.

In chapter 5 I present success semantics – roughly, the view that the truth conditions of beliefs can be derived from the success conditions of actions. By examining the problems faced by success semantics, I identify difficulties that have to be dealt with

Chapter 1

by my own project. I discuss two main problems: first, the possibility of ignored obstacles thwarting the success of actions, even if all the beliefs of the agent are accurate; second, the need to determine the conditions of success of actions.

In chapter 6 I revisit the second problem identified in chapter 5, that is, the problem of determining what makes an action count as successful or failed (in other words, the problem of determining the conditions of satisfaction of goals). I critically discuss two accounts of biological goals, namely evolutionary and organizational accounts, in order to see whether a satisfactory characterization of intentionality can be grounded in these accounts. The discussion of the problematic aspects of these approaches points towards the sort of proposal I will present myself (which is developed in chapter 8).

In chapter 7 I discuss how the possibility of ignored obstacles affects my own proposal. I claim that the fact that practical success can be thwarted by ignored obstacles highlights the defeasibility of practical reasoning. I argue that such defeasibility is a non-eliminable aspect of human representational practices and I explore some of its consequences (in particular, I suggest that defeasibility makes inferentialism an occasion-sensitive semantics).

In chapter 8 I put forward the account of success conditions that I endorse. On this account, the conditions of satisfaction of goals are ultimately determined by the evaluations of the agents subject to such goals. I argue that this approach manages to provide a satisfactory characterization of objective (even if fallible) representational practices, as long as the evaluative responses of agents tend, by and large, to track relevant environmental features.

In chapter 9 I present the main conclusions of the dissertation.

Chapter 2: A Practice-Based Approach to Intentionality

2.1 Introduction

In this chapter I present the main goal of the dissertation, which is to study the relations between the world and our speech and thought – in particular, to characterize representational practices from a pragmatist perspective. In the sort of pragmatist approach I have in mind, representation is studied by focusing on its role in human practices – intentional content is seen as conferred by use. I contrast this view with alternative representationalist proposals, which start by positing some words-world relation (e.g., informational theories or teleosemantics). I argue that, in order to incorporate a suitable notion of objective representation, pragmatist approaches must show how representational practices are constrained by the world. In this dissertation, I will propose that these constraints arise because representational practices involve practical interactions with the environment, and the outcomes of these interactions depend on the way the world is.

In section 2.2 I introduce the problem of representation and different ways in which it may be approached. In section 2.3 I present the sort of practice-based view of representation that I will explore in this dissertation. In section 2.4 I discuss the objective nature of representational practices and the constraints that such objectivity imposes on accounts of representation. Finally, in section 2.5 I examine the relations between practice-based accounts of representation and other alternatives to representationalism – in particular, expressivist accounts of language and thought.

2.2 The problem of representation

I will take it as an uncontroversial fact that we can talk and think about the world – we can describe how things are around us. So, the claim

(1) The table in my kitchen is red

describes the table in my kitchen. An assertion of (1) will be accurate or correct when the table in my kitchen is actually red and inaccurate otherwise. In the same way, my belief *that the table in my kitchen is red* will be correct or accurate when things are as described by such a belief.

I will say that when asserting or thinking things like (1) we engage in *representational practices*. As I noted at the start, I will assume that humans are able to engage in such practices. The question I want to address is how it is possible for us to do so – how we manage to engage in this sort of practice. In particular, I want to study those relations between the world and our speech and thought as a result of which such speech and thought can be said to be in the business of representing the world. I will focus on simple everyday cases of representation of the empirical world – that is, talk and thought about tables, chairs, apples and other similar middle-sized goods around us (I will only refer to more sophisticated or complex cases, like scientific representation, occasionally).

Assertions or thoughts representing the world will be said to have *intentional or representational content*. Intentionality, as I will understand the notion, is the feature of our speech and thought by virtue of which they can be said to be about – to be directed towards – the world. The intentional content of assertions or thoughts determines what they are about (what they represent). I will say that an account of intentional content is *representationalist* if it explains what it is to express or possess intentional content in terms of some posited word-world representational relation – such as reference, correspondence, extension or satisfaction. For instance, (1) has the content it has by virtue of representing the table in my kitchen as red. An assertion of (1) is correct when the way it represents the world to be corresponds with the way it actually is. Utterances of the sentence ‘The table in my kitchen is red’ express the claim (1) because the expression ‘The table in my kitchen’ refers to the table in my kitchen while ‘is red’ refers to (or denotes) the property of being red.¹

On the face of it, the possibility of representational practices seems unproblematic: it is a standard routine human practice to direct our attention towards things in the environment and to coordinate our expectations and behavior regarding such things.

¹ Alternatively, one could say that the predicate frame ‘x is red’ is satisfied when the variable ‘x’ is filled in by expressions whose extension is included in the extension of ‘red’.

However, from a different perspective, representation becomes enigmatic and difficult to account for. This is especially so if one endorses a broadly naturalistic point of view, according to which human beings and their practices are part of the natural world and as such should be describable, at least in principle, by the natural sciences. In the world described by the natural sciences, there do not seem to be intentional relations among things: objects in the world of physics do not represent each other, but merely interact causally (such interactions are not evaluated as accurate/inaccurate or correct/incorrect, as happens in the case of representations). How do intentional relations emerge in this natural world?

One option is to reject this puzzle as based on mistaken assumptions or a wrong way of looking at things: for instance, one can part ways with naturalistic worldviews, or just remain quietist and claim that representation is not a phenomenon that requires explanation (for a quietist view, see McDowell 1994). An alternative option is to try to naturalize representation. The aim here would be to account for intentional relations in terms of causal interactions. The content of speech acts or mental states describing the world would be characterized by appealing to some representational relation – that is, in terms of what they represent or what they refer to –, but these representational relations would be reduced to more naturalistically palatable notions.

Naturalistic reductionist accounts of representation have been popular in the last decades (especially since the 80's). There have been several projects in this direction: just to mention a few, Dretske's informational semantics (Dretske 1981, 1995), Fodor's asymmetric dependence account (Fodor 1987, 1990) or etiological semantics (Papineau 1987, 1993; Millikan 1984, 2005; Neander 1995) – also known as teleosemantics. Although some of these approaches are still quite popular today (in particular teleosemantics), it is fair to say that none of them can claim to be without problems and difficult challenges.

It is not my purpose in this dissertation to present conclusive or exhaustive objections against these reductionist projects (although I will refer to some of these projects at different points while discussing my own proposals, especially in chapter 6). I will not try to show, either, that quietist or non-naturalist positions are unsatisfactory. Instead, I will try to develop my own account of representational practices and their

emergence, remaining respectful towards a broadly naturalistic worldview. In the sort of approach I will follow, representation is studied by focusing on its role in human practices – intentional content is seen as conferred by use. What I intend to do is to study what characteristics a practice must have in order to count as representational. I will contrast this approach with representational proposals – which, as I said above, start by positing some representational relation in order to characterize intentionality. In the following section, I will introduce the sort of practice-based approach I will work with.

2.3 A pragmatist view of intentionality

2.3.1 Representationalism and pragmatism

Representationalist approaches characterize what it is to express or possess intentional content in terms of word-world representational relations (e.g. reference, extension or truth). The alternative approach I will present in this section goes back, at least, to the early pragmatists and Wittgenstein.² This alternative approach, which I will call *pragmatist*³ or *practice-based*, does not start from representational word-world relations like reference, but from the study of the use of language among the members of a community: how the use of expressions affects our behavior and our relations with other individuals and with the world. As the Wittgensteinian slogan goes, ‘the meaning of a word is its use in the language’.⁴ To be clear, the pragmatist approach I will explore does

²Besides figures such as Wittgenstein (1953), Sellars (1954) or Dummett (1981, 1993), contemporary authors endorsing this sort of practice-based approaches include Brandom (1994, 2000), Price (2011), Horwich (1998, 2005) or Gauker (2003)

³ The label ‘pragmatist’ is often used in the literature to refer to the sort of approach I have in mind (e.g. Price 2013; Brandom 2011; Blackburn 2010; Misak 2007). This usage is motivated by the focus of this approach on what agents *do* when engaging in representational practices, rather than on what representational relations are. However, it should be noted that the relations between this contemporary forms of pragmatism and the American classical pragmatists is not straightforward (see, for instance, Brandom 2011; Misak 2007).

⁴Wittgenstein says that this is so for ‘a *large* class of cases—though not for all – in which we employ the word “meaning”’ (PI 43).

not deny that words or beliefs stand in representational relations to the world; rather, the idea is that what it takes to express intentional content is not to be explained in terms of such relations, but in terms of use – in terms of the role played in our practices by words and states that stand in those representational relations.

Granted, 'use' is a broad and vague expression, and it may be possible to accommodate this slogan in most representationalist accounts, if one does not offer some specific elaboration. Before sketching the guiding lines of a pragmatist development of the meaning-as-use insight, let me provide some motivation for pursuing such pragmatist approach.

The representationalist strategy, remember, is to think of the expression of intentional content as resting on representational word-world relations, that is, relations connecting expressions – or the thoughts expressed by uses of expressions – and their referents, what they are representing. A first problem, discussed in the previous section is that it is not obvious how to explain, from a naturalist perspective, what it is to stand in such representational word-world relations. But, even if it is granted that there is some naturalist-friendly account of these representational relations, a second problem arises. For many kinds of vocabulary, it is not easy to see what the worldly relata could be in such word-world relations. There are certain areas of discourse in which it is difficult to find plausible referents for expressions which seem to be used in a representational way (at least referents that are acceptable from a naturalist perspective). This difficulty arises, for example, in moral, aesthetic or modal discourse, as well as in phenomenological or even causal talk. What objects and properties are we talking about when we make assertions about moral or aesthetic issues? It seems that a representationalist perspective requires that there exist objects and properties which are referred to by this talk (unless such talk is systematically failed or erroneous). Such a requirement may be problematic, in particular if one endorses a naturalistic point of view according to which the only objects and properties actually existing are those describable by science.

This is what Huw Price calls the 'placement problem' (Price 2011). Of course, one may always try to reduce aesthetic or moral properties to natural ones.⁵ If this reductive

⁵ See Boyd (1988, 2003a, 2003b), Jackson (1998).

project is not deemed promising, there are other possibilities for the naturalist. For instance, one may endorse a form of error theory and argue that assertions in those areas or discourse fail to find referents.⁶ Also, one could embrace a fictionalist position and claim that such areas of speech behave analogously to fictional discourse – in the sense that, in those discursive domains, the claims expressed by seemingly assertive utterances are not (or should not be) put forward as true, and accepting them does not require believing their contents.⁷ A further option is to endorse expressivism⁸ and argue that there is no genuine representational speech or thought in these domains of discourse: utterances and thoughts that look as if they were representing the world actually have some other function (e.g. express emotions or other non-descriptive attitudes of the speaker): by adopting this sort of expressivist view, the naturalist would get freed from the obligation to identify naturalistically-acceptable worldly referents for speech in the relevant area of discourse, since such speech would be taken to be non-representational after all. Alternatively, one may reject naturalism and become a realist about the existence of non-naturalistic properties⁹ (e.g. moral or aesthetic properties).

Whatever one thinks of the merits of these views, it counts against representationalism that it forces one to engage in such metaphysical debates, especially since there is no uncontroversial answer to them. To be clear, I am not arguing against the project of reducing moral or aesthetic properties to properties acceptable for

⁶ For instance, Mackie (1977).

⁷ For instance, fictionalism about mathematics is defended by Field (1980, 1989). Moral fictionalism is defended, among others by Joyce (2001), Hussai (2004), Kalderon (2005b) or Nolan, Restall and West (2005). For in depth discussion of fictionalist views, see the papers collected in Kalderon (2005a). Kalderon (2005a: 5-7) distinguishes hermeneutic and revolutionary forms of fictionalism: according to hermeneutic fictionalism, the target area of discourse actually works in analogous ways to fictional speech; in revolutionary versions of fictionalism, the target area of discourse *should be* treated in the way fictional speech is treated (revolutionary fictionalism would embrace error theory about the target area of discourse in its unrevised form – that is, when it is treated in a non-fictional way).

⁸ Earlier formulations of moral expressivism are found in Ayer (1936), Stevenson (1944, 1963) and Hare (1952). More recently, expressivism has been defended by Blackburn (1993, 1998) and Gibbard (1990, 2003); see also Chrisman (2007). Yalcin (2011) has developed an expressivist account of epistemic modals which explicitly respects compositionality (see also Schnieder 2010).

⁹ See Moore (1903); Shafer-Landau (2003).

naturalists: I remain neutral on the prospects of such a project. However, it *prima facie* counts against the appeal of a theory of intentionality if it commits one to this reductionist endeavor, on pain of either abandoning naturalism or embracing a form of error theory or fictionalism. It would certainly be attractive for the naturalist if there were a theory of intentionality that allowed her to account for representational discourse at face value without forcing her to undertake the demanding (and probably destined to failure) task of providing reductive naturalistic analysis for all the referents of representational speech (e.g. without forcing her to reduce moral properties to properties acceptable from a naturalistic viewpoint).

Of course, naturalism may be wrong. For the purposes of this thesis, however, I am assuming a broadly naturalistic perspective: my aim is to offer an account of representational practices that is compatible with this broadly naturalist perspective. It is also a possibility that error theory or fictionalism are after all correct for certain specific areas of speech; so, it seems that the representationalist may remain unmoved by the sort of considerations appealed by the placement problem: she could just endorse error theory or fictionalism when suitable (naturalist friendly) referents for some domain of discourse are not found. This position, while possible, becomes however less appealing if it has to be extended to large areas of our seemingly assertive discourse (e.g. assertive discourse about mental states, about causality or about meaning itself). All things equal, it is a welcome result if a theory of intentionality is compatible with the possibility that, by and large, when speakers treat their speech as representational and often correct, it is actually so. After all, our pre-theoretical intuitions as speakers about our linguistic practices are a *prima facie* valuable source of evidence when investigating human language and thought – and such intuitions do not seem to align with error theory or fictionalism, at least outside certain specific areas of speech. Certainly, the evidence provided by the pre-theoretical intuitions of language users can be defeated; but such evidence is nonetheless *prima facie* weighty, and one needs good reasons to overturn it. I do not think that representationalist approaches provide such good reasons, especially taking into account that, as I will try to show, there are alternative views that avoid these counterintuitive conclusions. A pervasive error theoretical perspective would involve a

radical revision of our appraisal of our discursive practices. If possible, a more conservative alternative is desirable.¹⁰

Thus, if one is a naturalist and wants to avoid error theory, the placement problem, as Price points out (2011), provides motivation for developing alternative accounts of the expression of intentional content – alternatives which do not rely on finding referents in the world for linguistic expressions. These considerations, of course, do not constitute a conclusive argument against representationalism, but only provide defeasible motivation to look for non-representationalist approaches. Nonetheless, added to the worries about naturalizing representational relations (discussed in the previous section), the placement problem offers good reasons for the naturalist to explore alternatives to representationalism.

Where should one look in order to find such alternatives to representationalism? It seems that an attractive candidate is the sort of pragmatist approach I have introduced above. If one explains what it is to express intentional content in terms of use, moral and aesthetic speech stops posing special difficulties.¹¹ One only has to describe the way expressions are used in those domains of discourse – the role that speech and thought in these domains plays in our lives and practices –, without needing to find naturalist-friendly worldly referents for the vocabulary used in those domains. Although it could turn out to be the case that specifying the use of moral or aesthetic vocabulary

¹⁰ Classical expressivist views, in which some seemingly descriptive domain of speech is characterized as actually not involving the expression of descriptive attitudes, are also problematic: in particular, it is notoriously difficult for expressivists to give an account of how the target speech works when interacting with logical connectives and in embedded contexts (see Schroeder (2008) for a detailed discussion); more generally, it is difficult to explain the uniform behavior of assertive speech in genuinely descriptive areas of discourse and in the allegedly non-descriptive domains targeted by expressivism (see Dreier 2004; also Chrisman 2008). In 2.5 I will discuss whether the pragmatist view I will present can be seen as a global form of expressivism (in which descriptive discourse itself is studied by considering the attitudes expressed by agents engaging in such discourse).

¹¹ Price (2011) and Chrisman (2008, 2010) try to show how the placement problem disappears when one substitutes a representationalist approach for a use-based one. Contemporary expressivists like Blackburn (1984, 1993, 2013) or Gibbard (1990, 2003) also tackle this problem from an anti-representationalist perspective, although they do not require that such anti-representational approach is extended to all areas of speech. There is a debate whether a stable solution of the placement problem demands a global anti-representationalist stance or whether a local one is enough (Price 2011, 2013; Blackburn 2013). I will return to this debate in section 2.5 below.

necessarily involved identifying moral or aesthetic properties with some worldly naturalistic property, the pragmatist is not committed to this being so.

Notice that the pragmatist point is not that representational semantic notions such as reference or extension cannot be legitimately appealed to when discussing the content expressed by utterances. Quite the contrary, the pragmatist will be happy to say that our speech represents the world, that the term 'table' is used to refer to tables or that the truth of what is claimed by uttering the sentence 'The table in my kitchen is red' is determined by facts about the table in my kitchen. Rather, the idea is that *what it is to express contents* is not ultimately explained by appeal to word-world relations underlying such representational semantic notions, but by studying the role in our practices of utterances expressing contents. Of course, the performance of assertions which can be said to refer to something will be often underlain by several relations between the relevant agents and the world (relations specifiable in non-intentional vocabulary). The pragmatist claim is that what it is for an assertion to refer to something should not be ultimately analyzed in terms of any of these word-world relations, but rather in terms of the broader role of referring assertions in the relevant discursive practice. This role will probably not be describable by focusing on any specific non-intentional word-world relation: there will not be a specific type of word-world relation that accounts systematically for the referring nature of assertions about tables and chairs, assertions in mathematics and also assertions about quarks and gluons (unless it is a relation specified in intentional terms, e.g. the relation of reference).

Likewise, the pragmatist may agree that the sentences 'James had the obligation to help Mary' and 'John has the belief that it will rain tomorrow' can be used, respectively, to describe James' obligations and to represent a situation in which John has the belief that it will rain tomorrow, without having to grant that there are naturalist-friendly worldly entities (i.e. entities that can be adequately specified in terms of the vocabulary of the natural sciences) that are referred to by uses of the words 'obligation' or 'belief'. It is enough if such sentences play a suitable role in our assertive practices. Of course, the naturalist will demand that these practices can be characterized in terms acceptable from a naturalist perspective. But offering this sort of global naturalist characterization of our practices is a different matter from finding reductive naturalistic analysis for each of the referents of the vocabulary involved in our representational

speech. Thus, even if one grants that it is hopeless to look for a naturalist reductive account of the property of being beautiful, unjust or an odd number, one may still think that it is plausible to give a general naturalist characterization of our representational practices (including our speech about ethics, mathematics, aesthetics, causality or mental states). The placement problem, therefore, does not seem to arise, at least in an obvious way, in the pragmatist approach I am considering.

Having provided some general motivation for a meaning-as-use, or practice-based approach to intentionality, let me now present in more detail the sort of practice-based framework I will work with.

2.3.2 Practice-based semantics

The core idea of practice-based perspectives, as advanced above, is to account for the expression of intentional content in terms of use: in general, to account for the possession of semantic features in terms of use (where semantic features are understood as those features characterizing the intentional content expressed by speech acts). Practice-based approaches can be seen as endorsing a certain view of the relations between semantics and pragmatics. Brandom condenses such pragmatist view in the slogan 'semantics must answer to pragmatics' (Brandom 1994: 83). In order to flesh out this slogan, Brandom proposes two basic principles constraining the relations between attributions of content and use – the relations between semantics and pragmatics (Brandom 2002, 2011; also MacFarlane 2010; see Dummett 1993 for a similar approach). I will follow Brandom and work with pragmatist approaches constrained by such principles.

The first of these principles is *methodological pragmatism* (Brandom 2002; Shapiro 2004):

Methodological pragmatism: semantic notions are introduced with the only purpose of accounting for the use speakers make of linguistic expressions.
(Brandom 1994; 2002)

This principle states that semantics is there to do ‘pragmatic explanatory work’ (Brandom, 1994: 144), this is to say, to explain or at least codify the pragmatic significance of linguistic moves¹² (‘pragmatic’ is used here to refer generally to aspects of linguistic use¹³).

On the other hand, it has to be specified how content is conferred by practices (Brandom 1994: 77). This requisite constitutes the second pragmatist principle, which Brandom calls *semantic pragmatism* (2002).

Semantic pragmatism: expressions mean what they do because of the way they are used.¹⁴

According to this second principle, all there is to know about the meaning of some sentence is how it is used in the relevant discursive practices (I will think of the meaning of linguistic expressions as assigning semantic contents to uses of those expressions). In this way, it is futile to attribute semantic differences to expressions with the same

¹² Dummett makes a similar claim: ‘corresponding to each different kind of force will be a different uniform pattern of derivation of the use of a sentence from its sense’ (Dummett 1981: 361).

¹³ The pragmatic significance of a linguistic move would be, in general, the effect that that move has in our practices. In other – narrower – common uses, pragmatics is related to the role of context in linguistic communication, or to the specification of the speaker's meaning, as opposed to conventional linguistic meaning; it has also been proposed that semantics deals with the information encoded in the linguistic aspects of an utterance, while pragmatics would focus on the additional information generated by the act of uttering a sentence (Bach 2002, 1999).

¹⁴ An example may help to see the difference between the two principles. Think of chemical substances. One first principle (analogous to methodological pragmatism) may demand that attributions of chemical structure contribute to explaining the properties of the substance in question. So, according to this first principle, saying that water is H₂O should contribute to explaining the behavior and properties of water (e.g. its being liquid in certain conditions, its freezing temperature, its density etc).

A principle analogous to semantic pragmatism (call it chemical pragmatism) would demand that attributions of chemical structure are justified or explained in terms of the properties of the substance. So, having a certain structure would just be a matter of possessing certain properties and deploying certain behavior.

pragmatic role (McFarlane 2010).¹⁵ Accordingly, what it is for a linguistic item to express – or for a mental attitude to possess – certain contents should be explained in terms of the role that such linguistic items – or mental attitudes – play in the relevant practices (e.g. how those linguistic items are used). Philosophical semantic theory – this is, semantics of natural languages – ‘incorporates an obligation to make the semantic notions it appeals to intelligible in terms of their pragmatic significance.’ (Brandom 1994: 145). In a strong version of pragmatism – which Brandom endorses, and so will I –, this pragmatic significance must be specified without resorting to semantic or intentional vocabulary (MacFarlane 2010). So, according to the sort of pragmatist approach I will follow, it is not enough to account for semantic notions by means of specifications of use which appeal to further semantic or intentional notions (e.g. it is not enough to explain the meaning of the term ‘table’ just by saying that it is used to *refer* to tables). Pragmatic explanations in terms of use must be intelligible without the involvement of semantic notions. While one may explain some particular semantic notion in terms of other semantic notions (say, one may try to account for the notion of representation in terms of the notion of satisfaction, or of reference), ultimately such explanations have to be grounded on semantics-free specifications of use.

It is important to note that both semantic pragmatist and representationalism (as I am characterizing it) are what I will call meta-semantic theses. That is, they are views about what it takes for some utterance (or mental state) to express semantic content: according to pragmatism, it is a matter of playing certain roles in a practice; according to representationalism, it is a matter of standing in certain word-world relations. Semantic theories, on the other hand, assign meanings to expressions, in order to characterize the content expressed by their use.

So, this is how I will understand the distinction between semantics and meta-semantics: a meta-semantic theory explains what makes something (say, an utterance or

¹⁵ In contrast, a non-pragmatist view of semantics may accept semantic differences that are not translated into differences of use. For instance, someone might want to attribute semantic differences to expressions which receive the same use but are associated, say, with different evolutionary histories – or with different neural patterns.

mental state) count as expressing contents with such and such semantic features – whereas semantic theories would characterized the semantic features of the contents expressed. Although different semantic theories may in principle be compatible with a given meta-semantic outlook, meta-semantics will impose constraints over the adequacy of semantic theories (for instance, a referentialist semantic theory, constructed with notions like reference or truth-condition, seems especially well-suited for a representationalist meta-semantic perspective).¹⁶

In the pragmatist meta-semantics I am presenting, the expression of intentional content is regarded as a matter of playing a certain role in a practice. An account of intentionality along these lines, therefore, must specify what sort of practices are discursive practices – that is, practices in which intentional contents are expressed. It must be explained, in other words, what kind of use a linguistic expression must receive in order to count expressing intentional contents. On the other hand, the pragmatist must provide a semantics (a specification of the contents expressed by utterances) which manages to satisfy the two principles stated above: attributions of semantic content must be suitably related to use.

In the following chapter I will characterize discursive practices and I will present a semantic framework which, I will argue, is naturally connected with use. To anticipate: I will model discursive practices as practices of giving and asking for reasons (Sellars 1954; Brandom 1994), and I will argue that inferentialist semantics (Brandom 1994) are naturally associated with such practices.¹⁷ Before doing so, however, I will devote the next section to discussing some features of objective representation which impose severe constraints on those practices that aspire to be counted as engaging in the representation of the world. In particular, I will explain why these constraints are difficult to satisfy by those wishing to describe representational practices in a naturalistic-friendly way – and I will sketch the strategies I will adopt in this dissertation in order to deal with those

¹⁶ Yalcin (2012) offers a clear account of the difference between semantic and meta-semantic (which he calls pragmatic) levels. See also Chrisman (2012):

¹⁷ As I will discuss in the next chapter, in principle semantics built on representational notions such as reference or extension could aspire to satisfy the pragmatist requirements I have outlined (see MacFarlane 2010). My claim will be that the connection with use is more easily established in inferentialist semantics – semantics built around semantic notions such as inference or entailment.

difficulties. The remaining of the chapter will focus on examining in some more detail the pragmatist project I have just sketched, situating it among related proposals.

2.4 Objectivity as answering to the world

It is a central feature of the notion of objective representation, as I will understand it, that representations (e.g. assertions or beliefs representing the world) can be assessed as correct or incorrect depending on whether things actually are as they are being represented. Thus, the way things are in the world determines the correctness of utterances or attitudes that purport to represent such things. For instance, an assertion of (1) will be correct when the table in my kitchen is actually red – that is, when the table in my kitchen is as described to be by the assertion. This idea is often expressed by saying that beliefs and assertions are correct only if they are true (only if they represent the world accurately).¹⁸ Although not all authors think so, I will assume that this idea of correctness is normative (for disagreement, see Hattiangadi 2006; Glüer and Wikforss 2009) – that is, the way the world is sets a normative standard in relation to which representations are assessed as appropriate or inappropriate. One may think of representational accuracy as an evaluative standard – so that inaccurate representations are in some sense bad or defective qua representations –, but for the purposes of this dissertation, I will think of it as establishing a normative deontic standard – that is, a standard bearing on what agents ought to, or may, do.¹⁹ Inaccurate representations are not only in a sense defective but, moreover, agents making them are doing something inappropriate. In the next chapter, I will present a model of discursive practices in which such normativity plays a prominent role.

¹⁸ It is sometimes also argued that, if some proposition is true, it is correct to assert it (see Whiting 2013 for discussion), but I will only commit myself here to the converse direction of entailment.

¹⁹ This is a standard view. Many authors in the literature consider that truth (accuracy) sets a norm determining what agents ought to, or may, believe or assert (some authors argue that such truth-norm is derived from a stronger knowledge-norm). For instance, see Brandom (1994), Whiting (2007, 2010), Williamson (2000), Boghossian (2008).

At any rate, even if you think that the standard of correctness associated with representational accuracy is not normative, it seems clear that representations – in order to count as such – must be subject to classification as correct or incorrect depending on their accuracy (depending on whether they get things right). This is an essential characteristic of representation.

I will say that representational practices are *objective* insofar as what is correct in those practices is determined by the way the things represented actually are, rather than by the assessing attitudes of agents evaluating such correctness. More generally, I will say that a standard of correctness is objective only if what is correct according to such a standard is not determined by the assessing attitudes of agents evaluating such correctness. Discursive practices, I submit, only count as representing the empirical world if they include objective standards of correctness – according to which correctness depends on how the world actually is.²⁰ I will assume that any satisfactory theory of representation has to be able to accommodate the objective character of representational discourse about the world.

The question the pragmatist has to answer, thus, is what sort of use a linguistic expression must receive in order to count as subject to such objective standards of correctness. Now, in the same way that the world described by physics is devoid of representational relations, it is also without standards of correctness (from the point of view of physics, events or objects do not count as correct or incorrect). Therefore, it is a challenge to account for the existence of such standards of correctness while remaining loyal to a naturalist worldview. So, from this naturalistic perspective, it will not do to posit some standard of correctness: it is necessary to explain how such standards may emerge in the sort of world described by the natural sciences.

²⁰ Perhaps there is a relevant notion of objectivity that applies to non-representational discourse. For instance, one could argue that mathematics are objective but not representational (although I take it that this last claim is controversial. It is not implausible that mathematical discourse is actually representational: it represents mathematical facts). Thus, I am not claiming that objective practices are always representational, but rather that discourse representing the empirical world around us is objective (it incorporates objective standards of correctness). Likewise, I leave open the possibility that there can be non-objective representational practices (although I won't commit myself to the intelligibility of such a possibility).

One popular strategy is to argue that standards of correctness are introduced by natural selection (Millikan 2005; Neander 1995). For the time being, I will not consider this option (I will discuss it critically in chapter 6). Instead, I will pursue a pragmatist approach, according to which things (e.g. performances) count as correct or incorrect by virtue of the way they are treated by agents within a practice. Thus, something becomes subject to some standard of correctness (in general, to normative standards) because of being regarded – and assessed – as subject to it. It is only because moves within a practice (e.g. an assertion) are treated by the participants as correct and incorrect that standards of correctness emerge in the practice. So, standards of correctness arise from the attitudes of agents – in particular, from their assessing attitudes²¹ (for this sort of approach, see Brandom 1994: ch. 1; also Haugeland 1982, 1998).

This kind of view is sometimes known as *pragmatism about norms*²² (Brandom 2011; Williams 2013) – a similar view may be constructed in relation to standards of correctness if one thinks that they are not always normative. The core of this view can be captured as follows (see Brandom 1994; 2011):

Pragmatism about norms: normative statuses are instituted by social practices.

Pragmatism about norms is often associated with Rorty (1979), according to whom social agreement is the source of all normative authority. Rorty endorsement of an extreme version of such pragmatism can be witnessed in the following quote:²³

As I see it, the whole point of pragmatism is to insist that we human beings are answerable only to one another. We are answerable only to those who answer to us—only to conversation partners. We are not responsible to the atoms or to God,

²¹ In the following chapter I will explain how assessing attitudes can be accounted for, ultimately, in terms of reinforcing behavior, which is naturalistically acceptable (see Haugeland 1982, 1998).

²² Again, the label ‘pragmatism’ is warranted – besides its being so used in the literature – because of the focus on what agents *do* when engaging in normative practices.

²³ Also: ‘no criterion that we have not created in the course of creating a practice, no standard of rationality that is not an appeal to such a criterion’(Rorty 1982: xlii).

at least not until they start conversing with us. (Rorty 2003; quoted in Stout 2007: 9)

Pragmatism about norms does well at accounting for the existence of normative standards: such standards are introduced by the evaluations made by agents. The problem with this approach is that it does not seem to leave room for *objective* normative standards – that is, standards fixed by the way things are, rather than by the evaluative attitudes of agents. As I will understand the notion of objectivity, a normative standard only counts as objective if being actually correct is not conflated with being regarded as correct from a particular assessing perspective. So, each subjective assessing perspective may be in principle mistaken about what is correct according to an objective normative standard. There is always conceptual room for distinguishing objective correctness and correctness as assessed from a particular perspective (see Brandom 1994: ch. 8).

The standards of correctness associated with representational practices are objective in this sense. What determines the accuracy (i.e. correctness) of a given representation is how things are in the world, not being regarded as accurate by some agent or group of agents. The accuracy of representations depends on the way the world actually is. In other words, representational practices answer to the way the world is (McDowell 1994): the world has authority, in some sense, over our representational practices.²⁴ So, the correctness of a given assertion or belief may overrun the assessing attitudes of all the participants of the practices – in the sense that all participants may be wrong about the correctness of an assertion or belief (Brandom 1994: ch. 8).

Rorty seems happy to abandon any aspiration to this sort of objectivity and argues openly that our discursive practices are not normatively constrained by the world – the world has no authority over them (see quote above). In this spirit, he defiantly proposes

²⁴ From now on, I will assume in my writing that the standards of correctness associated with representational accuracy are normative. Similar points could be made even if it is thought that such standards are socially instituted but non-normative.

to replace objectivity with ‘solidarity’²⁵ (Rorty 1991) and to assimilate truth to ‘what our peers, *ceteris paribus*, let us get away with saying’ (1979: 176).

These radical leanings have brought a bad reputation for pragmatism in many quarters.²⁶ However, there are authors who, while remaining sympathetic to a pragmatist approach, aspire to preserve a stronger notion of objective representation. These authors are sometimes gathered under the label ‘new pragmatism’ (see the essays collected in Misak 2007; also Levine 2008). New pragmatists aim at giving a pragmatist account of human language and thought which leaves room for the possibility of objective norms and representation. Insofar as I am endorsing a pragmatist perspective (in particular, pragmatism about norms) and I wish to account for objective representational practices, the goals of my project can be seen as aligned with those of new pragmatism.

Authors such as Robert Brandom (1994, 2008), Huw Price (2011), Michael Williams (2013), Jeffrey Stout (2007), Bjorn Ramberg (2000), Rebecca Kukla and Mark Lance (2009, 2014) or John Haugeland (1998) can be regarded as new pragmatists (in the sense I have just specified). Among the different new-pragmatist proposals, Robert Brandom’s inferentialist program (Brandom 1994, 2000, 2008) is usually acknowledged as one of the most developed and detailed analysis of representational practices. Because of this, I will base my own approach on Brandom’s pragmatist framework – although, of course, I will diverge from Brandom in several points, trying to overcome its shortcomings (see chapter 3, especially sections 3.5 and 3.7; also chapter 4, 6 and 8).

The challenge faced by authors in the orbit of new-pragmatism is to try to show how a notion of objective correctness – a notion of correctness determined by the way the world is, rather than by the attitudes of agents – can be made compatible with the pragmatist view that something is correct (or incorrect) by virtue of being treated as correct (or incorrect). If, as Brandom claims (1994: 296-297), normative statuses are instituted by the normative attitudes of agents, how can it be that there are objective

²⁵ ‘For pragmatists, the desire for objectivity is not the desire to escape the limitations of one’s community, but simply the desire for as much intersubjective agreement as possible, the desire to extend the reference of “us” as far as we can.’ (Rorty 1991: 22-23).

²⁶ Rorty’s views may be less radical than what these provocative quotes suggest. For a more nuanced presentation, see Stout (2007); also Levine (2008).

normative statuses that may overrun such normative statuses? In other words, if one endorses Rorty's insight that all normative authority has a social source (Levine 2010; Brandom 2011), it does not seem possible to allow for the world to have a say in what moves within a social practice are correct: the world cannot acquire authority over social practices, since such practices are being taken to be the ultimate depositaries of normative authority (Kukla and Lance 2014; Haugeland 1998).

A rough, and somewhat metaphoric, way of answering this challenge is by saying that, while it is us – agents with evaluative attitudes – who are the ultimate source of normative authority, we may decide to confer – by deferral – some such authority to the world. The participants of a normative practice may choose to delegate some of their authority to the world, so that certain features of the world will be taken to determine what counts as correct in that practice. Once we take such features of the world to determine correctness in this way, it is not up to us whether a certain move in our practices is actually correct. Correctness would stop being determined – at least directly – by the attitudes of agents. (This would be analogous to what happens when those with political authority in a country – say, the senate – decide to defer to the opinion of some experts in order to make some decision). The world would get a say in what counts as correct within the practice, but it would only acquire normative authority in an indirect way: it would be an authority by deferral, rather than original authority. It will be one of the purposes of this dissertation to spell out – in less metaphorical terms – the story just sketched.

One may be tempted to propose a quick solution to the challenge faced by pragmatist views. In this quick solution, one would just observe that human discursive practices are world-involving (Brandom 2008; Kukla and Lance 2014: 24-25), in the sense that such practices take place in the physical world and include interactions with the environment. Thus, it seems to follow straightforwardly that the world shapes and constrains human discursive practices.

It is true that there is a way in which discursive practices – practices involving language and thought – are obviously constrained by the world. So, for instance, the activity of constructing a building typically involves language and thought and, at the same time, it is clearly constrained by the materials employed, the sort of terrain the building rests on, the force of gravitation and so on. Building a house is an activity that

Chapter 2

may involve thought (e.g. one has to plan how to lay its foundations) but it also involves interactions with the world, such as grabbing bricks and mixing cement. However, this platitude should not lead one to dismiss too fast the challenge of objectivity. There are many normative practices that are not governed by objective standards and nevertheless involve this sort of worldly constraints. Therefore, the existence of such constraints does not guarantee the objectivity of the corresponding normativity practice. Think for instance of the norms of good manners in using fork and knife. Arguably, these norms are purely social: if a given community decides that it is good manners to eat holding the knife in a certain way, that will count as good manners in such a community. A whole community cannot be wrong about its norms of etiquette. Despite this lack of objectivity, the normative practice of eating with fork and knife is obviously constrained by the way the world is. So, the materials the forks are made of, together with the force of gravitation and other things, constrains the way in which forks can be employed by the participants of the practice. Likewise, the consistency of the food eaten determines the sort of cuts that will be produced. All this is true but, still, it cannot be said that the way the world is determines what counts as a (socially) appropriate way of using fork and knife in such a community (e.g. whether one ought to hold the knife with the right or left hand, or whether it is good manners to make cuts in a certain way).

I think that the idea that our representational practices answer to how things are by virtue of involving practical interactions with the environment is on the right track. However, such an idea has to be elaborated in order to get a satisfactory account of representation. What one needs to show is not only that the world constrains our practices but also that these constraints are such that the world gets to have a say in what counts as correct and incorrect within those practices (something it did not do in the example of norms about eating good manners). It is not enough that there is some friction between our representational practices and the world, it has to be the kind of friction which makes our practices answer to the way the world is (McDowell 1994) – the kind of friction which brings objectivity with it.

The aim of this dissertation is to try to argue that this sort of friction arises because of the practical engagement with the world involved in representational practices. In the pragmatist approach I will explore here, one would start by noticing that the outcomes of our practical interactions with the world are not (entirely) up to us, but

are determined by the way the world is. The success of our dealings with the environment does not only depend on our attitudes or opinions but also – and decisively – on how things are in the environment we are dealing with. So, whether the house I am building will actually stand the winter storms depends on whether the house is actually robust enough, on the strength of the storms etc – rather than on my thinking or wishing that it will succeed in doing so. Insofar as representational practices involve this kind of interactions with the world, one can decide to evaluate the moves within these practices in terms of their practical outcomes: one can look at such practical outcomes in order to assess some move in the practice as correct or incorrect. Given that practical outcomes (the success or failure of our practical dealings with the environment) are determined by the way the world is, our evaluations of the moves within the practices will also depend on the way the world is.

This is the project of this dissertation. Of course, there are many tricky points. For instance, the notion of success is, in the pragmatist approach I will follow, ultimately determined from within the practice: something is successful insofar as it is taken to be successful (that the house I built fell down immediately does not count as a failure unless the criterion for success in my practice of building houses is that they should stand). So, one would have to examine in what sense the evaluations of success made by agents can be said to be sensitive to the way the world is (this will be discussed in chapter 8). In general, one would have to show how this proposal manages to escape the pitfalls of similar pragmatist projects (in chapter 5 I will review some such projects).

In the rest of this dissertation, I will explore these issues and I will try to suggest ways for overcoming the sort of difficulties I have just pinpointed. In the remainder of this chapter, I will discuss the scope of the pragmatist approach to representation I am following and its relations to other non-representationalist views of thought and language.

2.5 Pragmatism, local and global

Earlier on, I appealed to the problem of finding referents for assertions involving moral, aesthetic or mathematical vocabulary – the placement problem – as one of the

motivations for looking for alternatives to representationalist accounts of intentional content. In the pragmatist meta-semantics I have presented as a promising alternative, discursive moves – i.e. assertions – are taken to express contents by virtue of their role in relevant practices (including their use to express attitudes), rather than by virtue of standing in representational word-world relations.

On the face of it, this pragmatist approach seems to be a natural ally of other alternatives to representationalist views, in particular an ally of modern versions of expressivism – also known as ‘quasi-realism’ (Blackburn 1993, 1998; Gibbard 1990, 2003). Expressivists agree with pragmatists in thinking that one should not try to account for the content of all assertions by positing representational relations to some referent in the world. Instead, they argue that the content of assertions in several domains of discourse – for instance, ethics, aesthetics or modal discourse – is better accounted for in terms of their role in the cognitive and communicative lives of agents. Rather than asking about the nature of the referents of moral – or aesthetic, or modal – concepts, one should ask about what agents do when they use such concepts in their thought or talk.

Thus, one might think that the sort of pragmatist meta-semantics I have introduced above would offer a general framework for the expressivist project (Price 2011; Chrisman 2008, 2010a, 2010b). However, some expressivists – notably Simon Blackburn (2010: 169-180, 2013) – have complained that such an unrestricted pragmatist program is too all-encompassing, and does not allow for the kind of fine-grained study of the distinctions among different vocabularies which characterizes the expressivist program. Modern expressivists like Blackburn argue that, behind the uniform surface of assertive discourse, there lie many different forms of thought and speech, with different functions and purposes. In particular, Blackburn and other expressivists want to be able to establish a contrast between genuinely descriptive discourse (paradigmatically, the discourse of the natural sciences, which describes the world) and other forms of assertive discourse – say, moral or aesthetic discourse – which have non-descriptive functions (e.g. they may be used to express non-descriptive attitudes). In this expressivist picture, vocabularies like the one of the natural sciences can be said to be genuinely descriptive because it is not problematic to find referents for them in the natural world – they do not incur in the kind of metaphysical embarrassment that lead to the placement problem.

Price calls the idea that there is a relevant contrast between genuinely descriptive and non-descriptive assertive speech the ‘bifurcation thesis’ (Price 2011b, 2014). Blackburn fears that if one adopts a pragmatist perspective over all sorts of assertive speech – that is, one characterizes, in general, content in terms of use rather than representational relations –, this contrast would be lost. This is why Blackburn (2010, 2013) is reluctant to endorse pragmatism across the board, and prefers to restrict his expressivist analysis to certain vocabularies, remaining a ‘local-expressivist’ – rather than a global one, as Price encourages him to do (Price 2014, 2013).

It is true that the sort of pragmatist view I have presented does not distinguish between domains of discourse in which the expression of content is explained in terms of use and domains in which the expression of content is explained in terms of representational relations: the expression of content in all areas of discourse is taken to be a matter of use. However, the fact that pragmatist approaches account for the expressions of intentional content, in general, in terms of use, does not mean that such approaches cannot make distinctions about the specific use of different vocabularies – in particular, distinctions about the fine-grained practical role of different sorts of assertive speech. So, as Huw Price argues (2007, 2011a, 2011b, 2013, 2014), the general account of assertion and intentional content offered by new-pragmatists such as Brandom (1994), can be combined with the expressivist’s detailed analysis of the several functions and purposes of different sorts of vocabulary. In this way, the pragmatist may acknowledge that assertions in different domains of discourse have different fine-grained functions – different effects for the relevant agents. Thus, it is possible for the pragmatist to keep a version of the contrast captured by the bifurcation thesis, although it will be a contrast not established in semantic or intentional terms – that is, it will not amount to a contrast between representational and non-representational forms of assertive discourse.

What is the specific form of use that distinguishes, then, the sort of assertive discourse involved in description of the natural world – as opposed to assertions about mathematics or aesthetics? Answering this question will be, in part, the objective of this dissertation. But a preliminary (and sketchy) reply can be attempted now. In the previous section I suggested that our thought and language about the natural world can be said to be objective – to answer to the way things are in the world – by virtue of involving practical interactions with the environment. These interactions are, at least partly,

underlain by causal relations with things in the natural world – say, with tables and chairs.²⁷ In particular, assertive speech about the natural world is directly constrained by the observational situation of agents and by the empirical experiences they undergo. So, if I am observing a red table, it is not correct for me (at least *prima facie*) to assert that the table is not red. Likewise, assertive speech about the natural world has consequences about the sort of outcomes that one can expect from our dealings with the environment. If I think that the tea is very hot, I know that I will burn if I drink it. If I assert that there is a table in the kitchen, you will expect to see a table when you enter the kitchen.

This kind of interaction with the world, underlain by causal relations with the environment, does not constrain in the same way other forms of assertive speech. So, observation does not play the same justificatory role in mathematics as in empirical sciences. If I say that there exists a natural number such that it is the square root of 4 – namely, 2 –, I do not commit myself to your being able to engage in relevant causal interaction with such a number: you will not expect to be able to manipulate number 2 physically, or to observe it when entering the kitchen.

So, the pragmatist can accept the existence of mathematical entities without having to renounce to naturalism. She just needs to point out that assertions about mathematics and about empirical facts have different roles in our practices – in particular, what it takes for a mathematical entity to exist is not the same that what it takes for a chair or a table (see Brandom 1994: ch. 5).

Maybe the rough sketch of the distinctive role of empirical speech I have just presented is not found altogether convincing by all (I hope it will have become more satisfactory by the end of this dissertation). In that case, alternative stories can be proposed.²⁸ But the main pragmatist point remains the same: the distinction between

²⁷ Note that the point of the pragmatist project is not, as Blackburn seems to think (2013), that one cannot mention, when providing stories about the use of certain vocabulary, the things about which such vocabulary is used to talk (say, chairs and tables). Rather, the idea is that these pragmatist stories cannot involve semantic word-world relations in order to explain intentional content (see Price 2014).

²⁸ For instance, Chrisman (2008, 2010a, 2010b) argues that what distinguishes descriptive and (non-descriptive) normative vocabulary is that the latter has a motivational-practical role that the former lacks (i.e. normative thoughts offer reasons for forming intentions to act).

what expressivists classify as descriptive and non-descriptive forms of assertive speech can be made in terms of use.

I have suggested that the interactions with the environment associated with empirical descriptions of the world might account for the objective representational dimension of such speech. What happens in the case of other forms of assertive speech in which causal relations with the world do not impose such constraints? One may wonder about the source of the constraints that make other forms of assertive discourse (e.g. discourse about mathematics or morals) objective— to the extent that they actually are objective – and the sort of discourse in which one may talk about facts, truth or knowledge. I think that at this point one may turn to the kind of stories provided by expressivism and quasi-realism about the uses received by different vocabularies and about how those uses explain the assertive and truth-apt character of such vocabularies (Blackburn 1993, 1998; Gibbard 1990, 2003).²⁹

In this dissertation I will not deal with these other forms of assertive discourse and their representational character. I will focus on empirical descriptions of the natural world – in particular, the sort of descriptive discourse typical of everyday language and thought about the way the world is around us. My target in this dissertation, therefore, will be the easiest, more mundane cases of representation, e.g. discourse describing the weather outside or the contents of one's fridge. It may be objected that this aim is too modest; surely, many interesting questions about representation only arise in less trivial cases (for instance, in discourse about theoretical, unobservable entities, about the distant past and future or about moral or aesthetic properties). However, pragmatist approaches have serious problems in accounting for the objective nature of representation even in simple, mundane cases; arguably, the most pressing difficulties for pragmatist approaches to representation are already manifest in these basic forms of descriptive speech. Thus, it will be a great step for pragmatist views if it is shown that a satisfactory account of objective representation may be achieved in relation to these easy cases.

What about the harder cases, then? As I said above, I will not have much to say about them here. The hope is that, once we get a clearer idea of how a pragmatist

²⁹ Although, of course, the pragmatist approach leaves the door open for alternative stories about the use of, say, moral or mathematical vocabularies.

account of representation may work in simple cases, we will be in a better position to address such less simple forms of representation. At any rate, note that most of these hard cases are hard for all theoretical approaches, and not only for pragmatism (it is a matter of debate how to deal with representation in moral or aesthetic discourse, or about theoretical entities). For what is worth, I think that pragmatism is actually better placed than alternative approaches to explain many of these non-obvious forms of representational discourse. Take theoretical, unobservable entities. Even if such entities are not directly observable, beliefs about them have more or less indirect practical ramifications. Thus, scientists take their beliefs about unobservable fundamental particles as their basis for designing experiments at CERN. The pragmatist can argue that, by virtue of such practical implications, discourse about theoretical entities has an objective, representational dimension.

So, my goal is not to give a general pragmatist account of all forms of representational discourse, but only of representation in mundane empirical discourse about the world around us (discourse about 'middle-sized dry goods'). This modest account, if successful, will pave the way for a more general pragmatist theory of representation – but such an ambitious project lies beyond the scope of this dissertation.

2.6 Conclusions

In this chapter I have presented the problem of representation, which will be the target of this dissertation. Briefly, the problem is to explain in what sense our language and thought can be said to represent the world – to be about things in the world.

It is common to account for representation in terms of word-world semantic relations (such as correspondence, truth or reference). However, the nature of such relations is obscure – in particular if one accepts a naturalistic worldview. Thus, I have proposed to follow an alternative pragmatist path, in which representation is accounted for by studying representational practices: the representational nature of linguistic expressions (and mental attitudes) is explained in terms of their use, that is, of their role in relevant practices.

In the second part of the chapter, I have discussed the main challenge of pragmatist accounts of representation. This challenge amounts to explaining how the

world may get to determine what counts as correct within discursive practices – in other words, how the world may acquire authority over our practices. I have suggested that the best way to deal with this challenge is to take into consideration the practical engagement with the world involved in discursive practices.

In the last section of the chapter, I have reviewed the relations of this pragmatist project to other anti-representationalist views, in particular the modern versions of expressivism endorsed by Blackburn or Gibbard. I have argued that the pragmatist approach I am following offers a suitable framework for developing the expressivist program.

In the following chapter I will present the semantic approach I will work with, namely, inferentialism (Brandom, 1994). I will argue that inferentialist semantics is easily connected with use and that, therefore, it is a natural option for the pragmatist looking for a practice-based semantics.

Chapter 3: Inferentialism as a Practice-Based Semantics

3.1 Introduction

The general aim of this work is to explore the representational dimension of human cognition and language from a pragmatist perspective – that is, from a perspective that focuses on the role of representation in our discursive practices. The strategy for doing so is to start by proposing a practice-based account of meaning – an account according to which expressing semantic content amounts to holding a certain position in a discursive practice. Representation is then studied by investigating the way in which the world constrains such practices.

In this chapter, I argue that inferentialism is the most natural framework for constructing a practice-based semantics (i.e. a semantics adapted to the sort of pragmatist meta-semantics introduced in chapter 2). First, I will introduce inferentialist semantics, that is, semantics in which the semantic content of claims is characterized by their inferential role (their inferential relations to other claims). Next, I will show how this semantics can be combined with the sort of pragmatist meta-semantics presented in the previous chapter, according to which expressing semantic content is a matter of playing a certain role in our practices and, moreover, attributions of semantic content must contribute to explaining use. Furthermore, I will examine certain features of inferentialist semantics that will be later relevant for my project.

So, the inferentialist framework I will present has two dimensions. First, there is an inferentialist semantic theory (which I will contrast with more mainstream alternative referentialist semantic theories). Second, there is a story connecting this semantics with a pragmatist meta-semantics according to which the expression for semantic content is a matter of use. This story will specify what agents need to do in order to count as expressing some semantic content characterized by certain inferential role. The key idea will be that the inferential profile of a claim can be naturally accounted for in terms of the role such a claim plays in practices of giving and asking for reasons, practices which, arguably, can be characterized without appeal to intentional or semantic notions. Practices of giving and asking for reasons will be described as normative practices in which certain performances are treated as correct or incorrect and in which the agents

making such performances acquire certain entitlements and commitments. Once this non-intentional characterization of practices of giving and asking for reasons is in place, it will be possible to meet the requirements of a pragmatist meta-semantics and explain in non-intentional terms what agents have to do in order to express semantic content: in a nutshell, an expression will count as having certain semantic content (characterized in terms of inferential role) when its use is governed by norms with a certain structure, more specifically a structure that mirrors the inferential profile of the content the expression counts as expressing. Inferential roles will be seen as codifying norms governing the use of expressions in practices of giving and asking for reasons.

My goal in this chapter, thus, is to discuss in detail how inferentialist semantics can be integrated in the framework of a pragmatist meta-semantics. The structure of the chapter is as follows. After discussing (in section 3.2) the pragmatist motivation for inferentialism, I explain in section 3.3 how semantic content and language use are related in Brandom's version of inferentialism: that is, I connect inferentialist semantics with a pragmatist meta-semantics. In section 3.4 I discuss the normative characterization of discursive practices provided by Brandom. Finally, in sections 3.5 to 3.7 I examine critically Brandom's theory and discuss some of the problems it faces, in particular its problems in accounting for objective representational practices. I end up by sketching how I intend to develop the inferentialist framework – in further chapters – so as to solve these problems and provide a satisfactory account of representation.

3.2 A pragmatist motivation for inferentialism

Inferentialist semantics characterizes the content of claims or propositions in terms of their inferential role – that is, in terms of the inferential entailments such claims are involved in. For instance, “The table is red” entails “The table is coloured”. According to inferentialist semantics, this entailment contributes to determining the content of such claims.

So, in inferentialism, the basic semantic notion – the notion used as a primitive in building semantic theories – is the notion of entailment or inferential connection. This contrasts with what I will call referentialist semantics, which take as their semantic

primitives notions like reference, truth condition, satisfaction or representation (for the label ‘referentialism’, see Williamson 2009; also Burgess forthcoming).

Referentialist semantics is often associated with representational meta-semantics – that is, approaches according to which the expression of semantic content is explained as a matter of standing in some representational word-world relation. The reason for this is that the semantic primitives of referentialism (e.g. reference or satisfaction) can be naturally understood in terms of such relations.

Inferentialist semantics, by contrast, lends itself to association with practice-based accounts of intentionality and meaning (i.e. with pragmatist or practice-based meta-semantics). This is so because inferential entailments between claims can be naturally understood in terms of the role those claims play (or ought to play) in reasoning or, more generally, in argumentative practices (see for instance, Brandom 1994: 91-92).

My main motivation for exploring inferentialism will be, precisely, that it offers a straightforward way of constructing a practice-based semantics, that is, a semantics that respects the two pragmatist meta-semantic principles introduced by Brandom (2002) and discussed in chapter 2: (i) *methodological pragmatism*, according to which semantic notions are introduced with the only purpose of explaining the use of language; and (ii) *semantic pragmatism*, which holds that expressions mean what they do because of the way they are used: expressing contents is taken to be a matter of receiving certain uses (specified in non-semantic terms).

I will not attempt to argue that referentialist semantics is unable to satisfy this kind of pragmatist meta-semantic aspirations. I think that such a possibility should be open; actually, one could think that a project along these lines is Davidson's, in which a semantics constructed from the notion of truth is employed to explain the behavior of agents (see MacFarlane 2010). Similarly, one can see success semantics (which will be discussed in chapter 5) as an attempt to combine a pragmatist meta-semantics with standard truth-conditional referential semantics. What I will try to show, instead, is that inferentialism is a natural framework for developing a practice-based account of meaning:

inferentialism is a natural choice for someone endorsing the sort of pragmatist meta-semantics discussed in chapter 2.³⁰

Perhaps it is possible to regard referentialism and inferentialism as complementary approaches, rather than competing alternatives. It could well be that referentialist and inferentialist semantics serve different explanatory purposes: which semantic framework is to be chosen would depend on our explanatory aims. So, referentialism is particularly useful if one wants to study certain systematic aspects of language use: for that purpose, the sort of compositional rules specified by referentialist semantics is illuminating. But, maybe, in order to give a practice-based account of semantic notions, it is more helpful to opt for an inferentialist framework: inferentialism would bring to light more clearly the relations between semantic notions and discursive practices (characterized in non-semantic ways). Each of these semantics approaches is better suited for different theoretical explanatory projects.

On the other hand, one may of course also choose a mixed approach which starts with both referential and inferential semantic primitives (McDowell 2005; Brandom 1994: 135 n90; 2008b; MacFarlane 2010). I will focus here, however, on a more radical perspective which eschews all representational concepts from the relevant semantic primitives. I will explore how far this radical approach can lead us in providing a practice-based analysis of the expression of content.

In the next section I will show how semantic features and discursive practices are connected in inferentialism. There are different ways in which an inferentialist framework can be fleshed out.³¹ Here, I will take Brandom's theory (1994) as my guide. The main

³⁰ For all I know, it could be possible that the same discursive practice could be explained both by semantics taking representationalist primitives, and by an inferentialist one. My claim is that it is easier to develop such use-based accounts of meaning if one starts from an inferentialist approach.

³¹ Gentzen's account of the semantics logical constants is an early example of inferentialism. One can also find inferentialist insights in Wittgenstein's work. Sellars, however, may be seen as the first author providing a general inferentialist semantic theory (1953). Conceptual role theories are other developments of inferentialism (Harman 1982; Block 1986). Paul Horwich's views (2005) are also, in many respects, inferentialist, although he diverges from Brandom's program in that he thinks that discursive practices are not inherently normative (I will discuss Brandom's normative description of language use in the next section).

reason for this is that Brandom's is probably the version of inferentialism which has been worked out in more depth, with a specific focus on developing a pragmatist approach to representation— which will be my principal interest in this dissertation. Of course, it is not necessary to follow Brandom in every detail, and I will not do so. There are certain aspects of his views which I will reject. In other cases, I will follow Brandom's specific proposals mainly with the aim of working with a concrete theory, rather than with an abstract schema – but the results I will arrive at will not depend on many of these specific choices.

3.3 Inferentialism as a practice-based semantics

3.3.1 Inference and practices of giving and asking for reasons

The basic tenet of inferentialist semantics is that the content of a claim is determined, at least partly, by its inferential connections with other claims – this is to say, by what other claims entail it and which other claims are entailed by it.

There is a straightforward connection between the inferential entailments a claim is involved in and the role played by such a claim in reasoning and, more generally, in argumentative practices of giving and asking for reasons. Saying that p entails q amounts to saying that having reasons to accept p provides reasons to accept q . For instance, the claim "The table is red" entails the claim "The table is coloured"; thus, one can offer reasons to accept the claim "The table is coloured" by giving reasons to accept the claim "The table is red".

Following Sellars, Brandom (1994: 90-91) exploits this connection in order to construe inferentialism as a practice-based semantics. The basic idea is that giving and asking for reasons is something that agents *do* – typically, something that agents do by means of linguistic moves. Thus, the content of claims would be determined by the place they occupy in a practice, namely a practice of giving and asking for reasons. However, this does not seem enough to satisfy full-fledged pragmatist meta-semantic requirements, because talk about reasons to accept a claim involves appealing to the semantic notion of content. Brandom's pragmatist meta-semantics does not only demand

that semantic features are made intelligible in terms of use; it demands also that this explanation in terms of use be free of intentional or semantic notions (Brandom 1994: 144-147; MacFarlane 2010).

Let me reconstruct Brandom's non-semantic account of discursive practices (I will present the way of reading Brandom's proposal which I think is more suitable for my purposes; I do not claim exegetical accuracy). According to Brandom, discursive practices essentially involve practices of giving and asking for reasons. Moreover, Brandom assumes that practices of giving and asking for reasons are inherently normative (1994: 7-18): within a given practice, certain moves are regarded as correct and others as incorrect. Agents engaging in these practices do not only make moves, but they also assess them. In particular, they provide reasons in order to vindicate the correctness of their discursive moves.

Normative notions, therefore, play an essential role in Brandom's description of discursive practices, whereas semantic notions are eschewed from such a description. In terms of such non-semantic (but normative) characterization of discursive practices, one can try to give a preliminary practice-based meta-semantics, that is a practice-based account of what it is to express semantic content (characterized inferentially). Very roughly, the idea is that the norms governing the correctness of discursive moves determine what contents such moves express. In this way, a linguistic expression *U* expresses a given content in virtue of the norms governing its use in relevant practices of giving and asking for reasons: when these norms mirror in relevant ways the inferential entailments involving the claim *that p*, *U* will count as expressing the claim *that p*. For instance, if it is correct for one to utter assertively the sentence 'The table is red', then it will also be correct to utter the sentence 'The table is coloured' and 'The table is not green'. This normative pattern of use mirrors the inferential links in which the claim *that the table is red* is involved, and because of this the sentence 'The table is red' counts as expressing such a claim. This account seems to fulfil the requisites of practice-based meta-semantics: on the one hand, attributions of content provide information about the use of expressions (in particular, such attributions provide information about the norms governing the use of expressions); furthermore, expressing content is taken to be a matter of being used in certain ways (in particular, being governed by certain norms).

In the next section I will discuss in more detail how the norms of a practice must be articulated in order to give rise to the expression of semantic content.

3.3.2 Commitments, entitlements and the expression of content

Brandom builds his analysis of discursive practices in terms of the deontic statuses of commitment and entitlement (Brandom, 1994, p. 159-161). Roughly, commitments – as Brandom understands them – are obligations, while entitlements are permissions. An agent is committed to doing A if it is inappropriate for her not to do it; on the other hand, someone is entitled to doing B if it is appropriate for her to do it. These two notions can be used to characterize the further normative status of incompatibility: two performances are incompatible if a commitment to one of them precludes entitlement to a commitment to the other one. Commitment and entitlement, understood in the way I have just presented, are clearly normative but non-semantic notions.

According to Brandom, by making moves in discursive practices, agents acquire and lose commitments and entitlements. In addition, Brandom argues that the normative structure of practices which count as expressing semantic content must be articulated in a complex way. In such complex practices, commitments and entitlements are not, in general, isolated, but related to each other (1994: 190-191). So, when an agent becomes committed or entitled to something, she also acquires further commitments and entitlements. A commitment to some performance brings commitments to other performances and, likewise, being entitled to doing something provides entitlements for making further moves. The norms governing the practice determine how different commitments and entitlements are related.

Brandom contends that assertion is the central move in discursive practices (1994: 167-168). Without moves playing the role of assertions, a practice cannot be counted as discursive or contentful. This is so because assertions are the essential tool in games of giving and asking for reasons. The paradigmatic way of giving reasons is by means of assertions and, according to the pragmatist meta-semantic I sketched above, discursive moves count as expressing semantic contents – characterized by their inferential profile – insofar as the role of such moves in practices of giving and asking for reasons mirrors the

inferential profile of the contents expressed. Without assertions, therefore, a practice would lack the sort of complex articulation which gives rise to the expression of semantic content.

I will explain this fundamental role of assertion below, but first, I will present Brandom's pragmatic characterization of assertive moves (Brandom 1994: 172-180; Wanderer 2008: 51-57). On the one hand, making an assertive move must imply acquiring certain commitments. On the other hand, an entitled performance of an assertive move will provide entitlement to the performance of further moves – in particular, to all those further moves one becomes committed to when making the assertion. Moreover, the commitments acquired by an agent when making an assertion must include the commitment to vindicating the entitlement to that assertive move (if such entitlement is suitably requested). When asserting, we become responsible for showing entitlement to the assertion made.

The conditions in the paragraph above characterize the intra-personal dimension of assertion; however, assertions also have an inter-personal side – derived from the fact that entitled assertions authorize the audience to take the speaker's word for what she asserted. An entitled performance of an assertive move counts as entitling speakers in the audience to making such an assertive move themselves. These speakers would acquire an entitlement *by deferral*: if challenged, they are entitled to transfer to the first speaker the responsibility to provide entitlement for the commitments associated with the move. By accepting this deferred responsibility, the original asserter authorizes her audience to make themselves the assertion in question.

This should be seen as a definition in normative terms of a technical notion called 'assertion'; according to what Wanderer calls Brandom's 'big bold conjecture' (Wanderer, 2008), the moves defined in this way would play the same role that assertions play in our discursive practices. On Brandom's view, the fundamental role of assertions in our practices is to be used to offer reasons – reasons for making other assertions, or for performing actions. In Brandom's normative framework, giving reasons for making some discursive move can be interpreted as vindicating one's entitlement to that move *by appealing to one's entitlement to some suitably related move*. Remember that I said above that an entitled assertion provides entitlement to further moves. Thus, by making an entitled assertion one can show one's entitlement to those further moves – one can

give reasons for making those moves. In this way, assertions are the standard move by means of which reasons for further moves are provided.

The central role of assertions in the expression of semantic content arises from this close connection between making assertions and giving reasons. In my reading of Brandom's meta-semantic theory, as discussed in section 3.3.2, some move can be said to express a certain claim p when the role of that move in the game of giving and asking for reasons mirrors the inferential relations involving the claim p . So, if the claim p entails the claim q , then an entitled assertion expressing the claim p will entitle the speaker to making a further assertion expressing the claim q . In general, when an entitlement to making some assertion provides entitlement to making some other assertion, those assertions can be seen as expressing claims or propositions which are inferentially connected.³² Brandom also characterizes a semantic counterpart to the normative status of incompatibility. When two claims p and q are semantically incompatible, the assertions expressing them are normatively incompatible – that is, making one of the assertions precludes entitlement to making the other one.

In Brandom's model, therefore, expressing contents is seen as a matter of playing a certain role in practices of giving and asking for reasons; furthermore, assertion is the central move of such practices – it is the paradigmatic way of offering reasons. This is why, according to Brandom, assertions are an essential component of discursive practices, so that there cannot be autonomous discursive practices without assertion.³³

³² I say in general because in some cases, an entitled assertion will provide entitlement to further assertions not because of the content of what was said (because of the semantic inferential connections involving the claim expressed), but because of the fact that the speaker has made such an utterance in that context. So, if a speaker makes an utterance asserting p , one will typically be entitled to asserting that the speaker has a voice (or some other mechanism for making utterances), which does not need to follow from the content expressed by the first assertion. Speakers would have an implicit grasp of the distinction between these sort of pragmatic transmissions of entitlement and semantic inferential links. According to Brandom, this distinction can be made explicit when speakers master suitable expressive tools – for example, conditionals: if the claim p is not asserted freely, but embedded as the antecedent of an asserted conditional, then only the semantic consequences of that claim may appear in the consequent.

³³ Of course, many other speech acts (questions, orders, proposals etc) can play very important roles in discursive practices. Brandom's claim is only that assertions are essential in order to understand how

Summing up, in Brandom's meta-semantics the inferential articulation of the semantic content associated with moves in the practice gets to be revealed *through the norms governing such moves*. In particular, what content is expressed by assertive moves is a matter of what commitments and entitlements are associated with making those moves: a move counts as expressing some content only insofar as the performance of that move brings certain commitments and entitlements to the agent – commitments and entitlements that mirror the inferential profile of the content expressed. More specifically, if some discursive move counts as an assertion of the claim p , then an entitled performance of such a move entitles, and commits one to accepting all further claims entailed by p . Moreover, the performance of such a move is entitled if one is entitled to accepting some other claim q from which p follows. This is, roughly, how the content expressed by an assertion is related to the entitlements and commitments associated to its performance. For instance, if an assertive utterance of the sentence 'The door is open' is to count as expressing the claim *that the door is open*, and an assertive utterance of the sentence 'The door is closed' is to count as expressing the claim *that the door is closed*, then entitlement to making the first assertion will provide entitlement to making the second assertion (given that the claim *that the door is open* entails the claim *that the door is not closed*). In the meta-semantic picture I am defending, therefore, the goodness of inferential entailments is understood in terms of transmission of entitlement: if entitlement to asserting some premises does not provide entitlement to asserting a certain conclusion, then the inferential link connecting those premises with that conclusion is not good (it is an invalid inference). Inferential entailment is a semantic notion, while the notion of entitlement belongs to the level of normative-pragmatics, to the normative characterization of practices (it is a permission to *do* something). The connection between the goodness of inferential entailments and the transmission of entitlement, thus, makes it possible to understand in terms of use what it is to assign semantic values: regarding some inferential entailment as good amounts to treating certain moves in the practice as correct, as entitled.

propositional contents get to be expressed. I will not discuss here the semantics of these other speech acts, and how they are related to assertions (see Kukla and Lance 2009).

In the next section I will discuss briefly what it is to be committed *to a claim* and how it can be reduced to being committed *to doing something*.

3.3.3 Commitment to a claim as a commitment to doing something

Brandom often says that when an agent makes an assertion expressing the claim p , she becomes *committed to the claim p* . However, it has been argued that this notion of commitment to a claim is obscure (MacFarlane 2010; Wanderer 2008). While we have an intuitive understanding of what it is to be *obliged to do something*, it is not clear how one can be *obliged to a claim*. Remember that Brandom's definition of commitment was about commitments to *doing something*; however, after introducing this definition he immediately goes on to talk about commitments and entitlements *to claims* – without explaining how this two kinds of commitments and entitlements are related (the swift transition between this two ways of talking can be witnessed in Brandom (1994: 159-160)).

Thus, it seems that the notion of commitment to a claim needs to be derived from the more familiar idea of commitments to doing things. There are different ways of doing so. For instance, it could be argued that if I am committed to p , I am committed – *prima facie* – to accepting p , and to being suitably disposed to assert p (for example, if I am questioned on the matter); in the same way, if I accepted an assertion expressing an incompatible claim, I would be doing something inappropriate. Also, if I am committed to p I ought to show entitlement to p if challenged. MacFarlane (2010: 91) thinks that this is circular – or very obscure –, since our clarification of a commitment to p would require understanding what is to show an entitlement to this same commitment. But maybe one could just say that such an entitlement is an entitlement to accepting or asserting p ; such an entitlement can be vindicated by doing things like (appropriately) asserting some other claim m that entails the claim p . Thus, I think that talk of commitments and entitlements to claims can be paraphrased in terms of talk about commitments and entitlements to doing things – although it is true that Brandom does not offer such paraphrases. From here on, whenever I use the notion of commitment or entitlement to a claim, it will be only as shorthand for this kind of paraphrase in terms of commitments to doing things.

3.3.4 Some features of the practices of giving and asking for reasons

Before concluding this section, I will mention two important points, (I will discuss them in detail at a later stage). First, according to Brandom (1994: ch. 4), practices of giving and asking for reasons may include both transitions from non-linguistic states (e.g., observational states) to discursive moves, and from discursive moves to non-linguistic performances (actions).³⁴ So, assertions may provide reasons not only for making further assertions, but also for the performance of non-linguistic actions; in the same way, the fact that a reliable agent is placed in certain observational circumstances may be taken as a reason for accepting some assertion. However, as Brandom stresses, assertions are the only move of the practices that both stands in need of reasons and may be used to offer reasons for other moves (1994: 167-168). I will discuss the connections of inferentialism with observation and action in chapter 4.

Second, speakers do not have to vindicate their entitlement to *all* of their commitments by appealing to some previous entitlement which confers entitlement to the commitment in question – that is, speakers do not have to provide reasons for all their discursive moves. There are many moves that will be in general seen as entitled *by default* (Brandom 1994: 204-206; Williams 2001). Challenges to the entitlement of these moves will not be regarded as appropriate unless there are good reasons that motivate them – and, at any rate, not all entitlements can be challenged at the same time. If discursive practices did not include this sort of default entitlements, it seems there would be a regress of deferred entitlements – there would be no end to the chain of transmitted entitlements. I will discuss these issues, in relation to the defeasibility of practices of giving and asking for reasons, in chapter 4 and in particular in chapters 7 and 8.

³⁴ Although he allows for the existence of purely inferential discursive practices that do not include connections with perceptions and action (see Brandom 1994: 167). Perception and action, however, play an essential role in practices that count as expressing empirical contents.

3.3.5 Summary

To sum up, according to Brandom's inferentialist meta-semantics, saying that some speakers express certain semantic contents with their discursive moves amounts to saying that the norms governing the performance of such moves mirror the inferential connections characterizing those contents. In particular, the norms determining the normative statuses acquired by speakers when making some moves would be isomorphic to the semantic inferential connections articulating the content those moves are taken to express. In other words, semantic inferential relations *codify* the norms governing the use of the discursive moves which express such contents (Brandom 1994: 627). These norms regulate not only linguistic transitions some move is involved in (i.e., what other linguistic moves it entitles and commits us to make), but also the non-discursive observational circumstances that may entitle the move, and its proper practical consequences. In this way, inferentialism connects meaning with a general model of discursive practices which covers most of the areas of language use to which use-based theoreticians typically appeal. This is why inferentialism can be seen as constituting the best framework for building meaning-as-use theories (so, meaning-as-use theorists such as Horwich (2011) have adopted some version of inferentialism). Of course, this general model will be rather schematic: a fine grained analysis of each particular area of discourse will require different developments of the inferentialist model, specially tailored for those specific domains

One word of warning. The notion of inference can be used in two ways: to refer to the act of inferring – an action carried by some agent – and also to refer to semantic relations among claims or propositions – as when we talk about the inferential entailment between two propositions, or the inferential structure of some argument. As a result of this dual dimension, focusing on the notion of inference allows one to highlight the connections between meaning and use – inference is a doing, and also constitutes meaning. But if one indulges in slippages between these two senses, there is the risk of blurring the distinction between the level of semantics and the level of use. Sure, these two levels are – in a pragmatist approach – related, but this does not mean that one should conflate them. For instance, the claim “The table is red” entails the claim “The table is coloured” – this is a semantic inferential connection. However, the performance of an utterance expressing the claim “The table is red” does not *entail* (in a semantic

sense) the performance of an utterance expressing the second claim. According to the norms of the practice, if one is entitled to make the first utterance, then one is entitled to make the second one – this is a normative-pragmatic connection between the appropriateness of making different moves, not a semantic relation.

I would say that a related distinction applies for the case of reasons. If the claim p entails the claim q , it is the fact that p – rather than the utterance of an assertion expressing p – which is a reason for some agent to accept q . The utterance of an assertion expressing p may, if properly entitled, be used to *provide* reasons for accepting q (but the utterance itself does not constitute those reasons). It has to be said that Brandom is not always as clear as would be desirable in keeping these levels distinct.³⁵ In order to avoid confusions, unless I explicitly say otherwise I will reserve the notion of inferential entailments to talk about the articulation of semantic content – rather than about the act of inferring.

Before discussing some further semantic features of Brandom's inferentialism, I will devote next section to explore in more detail the normative structure of discursive practices, and how this normative structure emerges from the doings of the participants of the practice.

3.4 Pragmatism and the normative dimension of discursive practices: normative statuses and normative attitudes

The main motivation for following the pragmatist route and explaining semantic features in terms of use is the hope that this will make such features less mysterious or obscure. Nevertheless, it is not clear that our grasp of normative concepts like commitment is better than our grasp of semantic or intentional ones (Rosen 2001; Laurier 2009). Brandom is aware of this, so he proposes to elucidate the nature of normative statuses by adopting what he calls 'phenomenalism about norms' (1994: 25, 334) or, more recently,

³⁵ For instance, at some points he talks of inferential relations as relations among commitments or entitlements (1994: 168-169), but also as relations among sentences (1994: 189) or among propositions (1994: 91).

'pragmatism about norms' (2011) – as already pointed out in chapter 2. The idea is to understand normative statuses by focusing on what agents do when they take someone to possess such statuses. In Brandom's words: 'Normative statuses are domesticated by being understood in terms of normative attitudes' (1994: 626). The fact that some move is inappropriate is shown in the attitudes of users, who – if they act properly – will (implicitly) treat it as inappropriate. Normative statuses – commitments and entitlements – only play a role in the practice through the attitudes of speakers (1994: 133-134, 161).

Brandom distinguishes two basic normative attitudes. On the one hand, agents may *attribute* normative statuses – commitments or entitlements. If I attribute the commitment to doing A to some speaker, and the speaker does not fulfil such a commitment – does not do A –, I will assess her performance as inappropriate (1994: 165-166). On the other hand, agents may *acknowledge* commitments or entitlements – that is, they may treat themselves as committed or entitled to something. Acknowledging a commitment or entitlement implies attributing it to oneself (the typical way of acknowledging one's commitment to accepting some claim is by making an assertion expressing such a claim).

Once the notion of attributing is in play, *undertaking* a commitment or entitlement can be described as doing something that makes it appropriate for others to attribute to the agent that commitment or entitlement. Note that, according to Brandom, undertaking a commitment does not always require acknowledging it: an agent can undertake some commitment because of what follows from what she said, even if she is not aware of these consequences. Brandom calls this sort of undertaken but not necessarily acknowledged commitment 'consequential commitments' (1994: 193-198).³⁶

So, according to Brandom's account of discursive practices, participants of such practices do not only make moves – produce performances –, but they also assess the

³⁶ Brandom takes an agent to believe some claim when she is committed to such a claim. However, he points out that this characterization can be read in two ways (1994: 194-196) On the one hand, one may take into account those commitments that the agent acknowledges; on the other hand, our ascriptions of beliefs may be concerned with the commitments the agent has actually undertaken, even if she does not necessarily acknowledge them – that is, about what she *ought to think* (given what she endorses). Brandom thinks that attributions of beliefs actually show this duality.

performances of others – they have normative attitudes towards those performances. Being a competent speaker means being able to track the different commitments and entitlements acquired by the relevant speakers – including oneself. Brandom calls the set of commitments and entitlements of an agent her 'deontic score'. *Scorekeeping* would be the activity of keeping track of this score (1994: 181-187). Actually, the norms that govern the acquisition of commitments and entitlements in the practice should be seen as *norms of scorekeeping* (MacFarlane 2010). According to Brandom, in complexly articulated practices, acquiring a commitment brings *automatically* the acquisition of further commitments. For instance, accepting the claim "France is bigger than Germany" commits one to accepting the claim "France is smaller than Germany" (Laurier 2009; Brandom 1994: 190-191). It is not that there is a norm according to which the speaker does something incorrect if she does not acquire that second commitment: she cannot help acquiring it. The norm relevant here is a norm about scorekeeping: a scorekeeper *ought to attribute* that second commitment to any speaker to whom the first one is also attributed. The norms governing discursive practices, therefore, are norms about the appropriate normative attitudes of agents (Brandom 1994: 182-186).

On Brandom's view, therefore, normative statuses can be understood in terms of normative attitudes. Brandom argues that attitudes, in turn, will be typically revealed by the sanctioning behavior of individuals (Brandom 1994: 34-35). If my performance is taken to be inappropriate, other agents will be disposed to sanction it accordingly – although one can always question if these sanctions are themselves appropriate. Some sanctions may be specifiable in non-normative terms – i.e., as positive and negative reinforcement. However, Brandom is keen on allowing for the possibility of sanctions specifiable only in normative terms, which he calls 'internal' sanctions; these sanctions would consist of alterations of the normative status of the agent sanctioned (1994: 44-46). For instance, a positive 'internal' sanction may amount to conceding further entitlements. Another example of an internal sanction is to exclude the infractor from the normative practice in question: for instance, someone who persists in not fulfilling her promises will end up by not being counted as making promises at all – the other speakers will not rely in her word, and will not form the correspondent expectations about her behavior (Brandom 1994: 180).

Of course, there are still many difficulties with Brandom's project of explaining normative statuses in terms of the attitudes of agents – in particular, the problem of accounting for objective normative statuses, which I will discuss in a later section. But, if one thinks that this project is on the right track, then Brandom's proposal offers a promising framework for connecting the semantic features of a language with what the speakers of that language *do*.

In the next section I will review some aspects of the kind of semantic contents that get to be expressed by the normative practices described by Brandom's inferentialism.

3.5 The semantics of inferentialism

In the past sections I have introduced an inferentialist framework, constituted by an inferentialist semantics, according to which semantic content is characterized by the inferential relations among claims, and a pragmatist meta-semantics, according to which the inferential relations characterizing contents mirror the normative connections among the moves expressing such contents. In this section I will explore the nature of the inferential links that, in the inferentialist semantics I am presenting, characterize content – taking into account the kind of normative practices that express those contents.

3.5.1 Material validity

The first point that should be noted is that the relevant inferential links will not, in general, constitute formally valid arguments. The reason for this is obvious. Formally valid arguments are good in virtue of their logical form, independently of the specific contents of the non-logical concepts involved. Therefore, this sort of inference will not allow one to characterize the content of those non-logical concepts. Inferentialists like Brandom choose instead *material goodness* as the basic notion of inferential goodness (Sellars 1953; Brandom 1994: 97-105). Materially good inferences are usually said to be good in virtue of the meaning of the claims involved. However, since, in an inferentialist approach, the meaning of a claim is determined, precisely, by the good material inferences it is involved in, this seems to be circular (Brandom 1994: 134). It is better to

take material goodness as a primitive notion of inferential goodness, which will then be used to characterize the contents of claims (at the pragmatic level, this primitive notion of inferential goodness has its counterpart in the normative notion of an appropriate transition between moves). Examples of material inferences are the inference connecting “Mary is taller than John” and “John is shorter than Mary”, or the one linking “The water is boiling” and “The water is very hot”.

Formally valid inferences can be derived as particular cases of materially good inferences: they would be the subset of materially good inferences which remain good under substitutions if we maintain invariant certain fixed vocabulary – namely, logical vocabulary. I do not need to commit myself here to any specific view of logical vocabulary and its functions, but I will briefly present Brandom’s proposal as an interesting example of how an account of logic may be developed from an inferentialist perspective. Brandom thinks that logical concepts – in a broad sense – can be identified as those concepts which have an expressive role, that is, which are used to make explicit aspects of the inferential articulation of the semantic content expressed in the practice (1994: 104-116). So, for instance, conditionals would be used to codify explicitly inferential connections among claims. In this way, by asserting the relevant conditional, speakers may express explicitly their endorsement of a given inferential link – an endorsement which, in the absence of conditionals, would remain implicit in the speaker's practical treatment of that inferential link as good. Likewise, conditionals allow agents to criticize and give reasons for the acceptance of an inferential connection – and to explore the unforeseen consequences of such an acceptance. This is why Brandom thinks of logic as a tool of rational criticism, which allows us to illuminate, amend and improve our discursive practices.

Many referentialist semantic notions have this sort of expressive function in an inferentialist framework – which makes them, in Brandom’s view, count as logical in an extended sense. The concept of truth is a significant example. Brandom, like most meaning-as-use theoreticians, endorses a version of deflationism about truth (1994: 299-305).³⁷ According to Brandom, the content of “It is true that p ” is the same as the content

³⁷ Brandom endorses a prosentential version of deflationism (1994: 301-305), according to which ‘...is true’ is a prosentence-forming operator. This means that when that operator is combined with some construction which specifies some claim or set of claims (like ‘What he said’ or ‘That it is raining’), a

of “*p*” (i.e. the inferential role of both claims is the same); in the same way, the content of “What the student said is true” shares the content of whatever the student said. The main role of the concept of truth is to function as a disquotational tool and to allow speakers to express the content of certain generalizations (as in “Everything the student said is true”), or of claims that can be described or nominalized but not expressed directly by the speaker (as in “The complex theorem proposed by the Russian professor is true”).

As required by pragmatist meta-semantics, the analysis of what it is for a claim to be true is not ultimately grounded in any appeal to word-world representational relations.³⁸ Instead, the notion of truth is characterized in terms of its relations to other semantic notions (i.e. inferential entailment); in turn, what it is to possess or express these further semantic features is accounted for in terms of use.

The inferentialist, thus, is ready to accept biconditionals like the following:

(1) ‘The table is red’ is true iff the property of being red is instantiated by the table.

(1) only says that the content expressed by assertions of the quoted sentence is the same as the content of the claim at the right-hand side of the biconditional: one is entitled to utter assertively ‘The table is red’ whenever one is entitled to assert that the property of being red is instantiated by the table.

Therefore, inferentialism is in a position to accept most of the results of formal referential semantics, even if such results would not be taken to offer an account of what it is to express semantic content. Instead, formal referentialist semantics would help to

prosentence is formed ('What he said is true'). The content expressed by that prosentence would be the content of the claims specified anaphorically by whatever is combined with the operator ‘...is true’. The details of this specific version of deflationism, however, are not important for my purposes.

³⁸ Brandom develops an anaphorical account of reference along similar lines (1994: 305-333; also Burgess forthcoming). However, discussing this would require explaining Brandom's approach to sub-sentential semantics, which is not directly relevant for my purposes here.

articulate explicitly the relations among the meanings of different expressions (Brandom 1994: 143-45; Horwich 2008).

3.5.2 Inferential non-monotonicity

In an inferentialist semantics the contents of claims are determined by the *materially* good inferences such claims take part in. But, what are the characteristics of these material, content-constituting inferences and what sort of semantics arises from them? One first point – which will be important later – is that many material inferences are defeasible, i.e. non-monotonic. The goodness of a non-monotonic inference may be defeated by the introduction of additional premises. That is, when new premises are added, it may be the case that a conclusion that followed from the former set of premises does not follow from the new expanded set. This is what happens in inferences like the one connecting “The water is boiling” with “The water is very hot”. If we add the premise “The pressure is extremely low”, the former conclusion stops being entailed. Another example is the inference from “The brake pedal is being pressed” to “The car will slow down”: if the brakes are broken, or the road is too slippery (or any of many other defeating possibilities takes place), the car will not decelerate.

Although the inferences in classical logic, mathematics and maybe certain areas of fundamental physics are typically monotonic, it seems that many of the material inferences involved in human reasoning and communication are not. It is in principle possible to stipulate that content is determined only by monotonic inferences, but Brandom explicitly chooses to include also non-monotonic ones (Brandom 2000, 2008) – and it is too be expected that an inferentialist theory that intends to account for the content of all claims will have to take into account this kind of inferences. The non-monotonicity of material inferences will crucially shape representational practices: the possibility of indefinitely many obstacles thwarting the success of our actions is one manifestation of such non-monotonicity. I will discuss it at length in chapter 7 (and also in chapter 4).

3.5.3 Commitment-preserving and entitlement-preserving inferences

So, one first classification of material inferences is between monotonic and non-monotonic ones. Can inferentialist semantics be illuminated by proposing further classifications of material inferences? Brandom thinks so. He tries to distinguish different kinds of semantic entailment by attending to their pragmatic manifestations – more specifically, by attending to the way in which the deontic score of agents is updated when they become committed or entitled to the claims involved in those inferences (1994: 168-169). On the one hand, semantic entailment will generally be *entitlement-preserving*. This means that, if an agent is entitled to asserting or accepting the premises of some inference, she will also count as entitled to asserting or accepting the conclusion. On the other hand, according to Brandom some inferences are *commitment-preserving*, or *committive*. In this case, if an agent is committed to the premises of the inference, she will also count as committed to the conclusions. A way of unpacking this formulation in terms of commitments to doing something is the following: an agent asserting, or otherwise endorsing the premises becomes committed to being suitably disposed to endorse or assert the conclusions (for example, when questioned about them); also, the agent would not be entitled to make an assertion incompatible with the conclusions. An example is the inference from “The table is red” to “The table is coloured”. Brandom argues that not all inferences are commitment-preserving. Some inferences, which he calls *permissive*, are only entitlement-preserving. The reason for this is that, in these inferences, ‘the possibility of entitlements incompatible with the conclusion is left open’ (1994: 169; also, 189). So, agents reasoning in accordance with good permissive inferences may be properly entitled to incompatible conclusions. Brandom thinks that inductive inferences are of this kind. Finally, Brandom also identifies a third genus, *incompatibility-entailments*, which are inferences in which everything incompatible with a commitment to the conclusions is also incompatible with a commitment to the premises (Brandom 2008). So, if some claim is incompatible with the claim “This animal is a mammal”, it is also incompatible with the claim “This animal is a dog”. Therefore, “This animal is a dog” incompatibility-entails “This animal is a mammal”. Brandom stipulates that the relation of incompatibility is non-defeasible (if p is incompatible with q , it is also incompatible with q and r), so incompatibility-entailments will just cover counter-factually robust (i.e., monotonic) inferences.

I think that Brandom's tripartite distinction is misguided. In particular, I do not think that there is any interesting use for the idea of good entitlement-preserving inferences entitling incompatible conclusions – at least, outside the realm of practical reason, as I will explain shortly.

One first problem is that many inductive inferences seem to be clearly commitment-preserving, contrary to what Brandom thinks. Take the inference from “The water is boiling” to “The water is very hot”. In standard circumstances, if I am committed to accepting the premises, I will be seen as committed to accepting the conclusions. If I threw boiling water to the face of my enemy, I could not excuse myself in a trial by saying that I knew that the water was boiling, so I was entitled to conclude that it was very hot, but that I was *not compelled* to commit myself to accepting such a conclusion. I think that something similar happens with all good inferences accepted on inductive grounds.³⁹

But, leaving aside this question, my main problem is with Brandom's contention that good (and sound) permissive inferences may entitle an agent to incompatible conclusions – more specifically, with his claim that this form of reasoning is pervasive in theoretical reasoning, and gives rise to a special sort of inferential entailment. As I see it, only in practical reasoning does it seem possible to find patterns of reasoning entitling incompatible conclusions – and which are, therefore, entitlement-preserving but not commitment-preserving. For instance, if my goal is to remain dry, and I know that it is raining, I have good reasons to undertake any of several (incompatible) courses of action: I can either open my umbrella, or go into the car, or enter the restaurant overhead.

One first thing to say about this sort of examples is that it is far from clear that there is some sort of 'practical' semantic inferential connection underlying these instances of practical reasoning. Note that this piece of reasoning ends in an action: what would be the conclusion in the corresponding semantic inferential link?⁴⁰

³⁹ Maybe some of these inferences should be interpreted as probabilistic inferences, which offer support for a partial belief in the conclusions – for a certain degree of acceptance of such conclusions. However, Brandom is not defending this possibility, but rather the idea that, while agents are entitled to a full acceptance of the conclusion, they are not compelled to accept it.

⁴⁰ I will discuss this issue in the next chapter, in section 4. 3.5. A possibility is to argue is that the conclusion of the inference is a normative claim, such as “I ought to/may open my umbrella”

Anyway, even if there were some sort of permissive (entitlement-preserving) semantic entailment associated with practical reasoning, Brandom's position would still be problematic. He thinks that this kind of example is all around, and can also be found outside practical reasoning – after all, it would include all inductive inferences. However, it is very doubtful that this is so. As I see it, in the theoretical domain (that is, in our descriptions of the world), agents cannot be *properly* entitled to incompatible claims – at least outside areas of discourse with relativist semantics. There is a simple reason for this. If one wants to preserve some degree of objectivity in our descriptions of the world, as Brandom does, it is necessary that the correctness of those descriptions depends on how things are in the world. Thus, an agent cannot be properly entitled to accept a false claim – i.e., there must be something incorrect in accepting a false claim. And, in two incompatible claims, one is bound to be false (again, at least outside relativist semantics). Thus, two incompatible claims cannot be properly entitled.⁴¹ I think this is a regulative principle of our practices of describing the world. Whenever an agent finds equally strong considerations supporting incompatible descriptions of the world, she cannot take herself to be entitled to any of those descriptions. Dancy criticizes Brandom's views in a similar spirit: 'In theoretical reason there is no analogue of the kind of plural entitlement that we find in practical reason.' (Dancy 2004: 62).

Brandom seems to think that non-monotonicity is the source of permissive inferential connections (1994: 169). The idea is that, when a defeating premise is added, the agent may become entitled to accept a new conclusion that is incompatible with the conclusions available before the introduction of the new premise. In Brandom's example: an entitlement to the premise "The match is struck" entitles one to the conclusion "The match will light"; but if one is also entitled to "The match is wet", then one will become entitled to incompatible conclusion "The match won't light".

Does non-monotonic reasoning really offer the possibility of entitlement to incompatible conclusions? It does not seem so. When the defeating premise is added the agent would lose her entitlement to the first conclusion ("The match will light"). So at no point is the agent entitled to *both* incompatible conclusions. Gaining entitlement to one

⁴¹ Although agents may be epistemically blameless when making a false claim, for which they are not properly entitled. Perhaps there seemed to be very good reasons supporting such a claim.

of them means losing entitlement to the other. It seems, therefore, that Brandom is interpreting wrongly the dynamics of defeasible reasoning.

Brandom actually rejects explicitly the strong notion of entitlement I have been appealing to – a notion that precludes proper entitlement to a false claim (Brandom 1995; 2000: 192-204). His motivation for this rejection is that, as he sees it, a notion of entitlement subject to such a stringent requirements does not play any role in a social practices of giving and asking for reasons; this sort of entitlement would be of no use in characterizing what agents do in those practices – and, besides, it would turn inferentialism into a standard truth-conditional semantics. Presumably, the worry is that scorekeepers would never be in a position to attribute entitlements to speakers (Brandom 2000: 201) – since, no matter how justified the speaker appears to be, it is always possible that what she says is nevertheless false.

I think, however, that this strong notion of entitlement has a perfectly natural place in a pragmatist account of discursive practices – and that, actually, this strong notion offers the best hope for developing a pragmatist approach which leaves some room for objectivity. The crucial point is to see attributions of entitlement as *defeasible*. Entitlements of this kind would be attributable to speakers when their views are sufficiently well grounded; however, if good reasons against such views appear, the attribution of entitlement will be revoked – the attribution would have turned out to be actually inappropriate. Attributions of entitlement would always be open to revision, and will be updated in the light of new evidence. This is the notion of entitlement I will work with in this dissertation (particularly in chapters 7 and 8), with the aim of accounting for the open-ended correctable nature of objective discourse.

In conclusion, I think that Brandom's characterization of the varieties of material inference is ill motivated. A careful examination of defeasible reasoning shows that non-monotonicity does not require allowing for the possibility of entitlement to incompatible claims. Furthermore, this possibility is undesirable if one wants discursive practices to accommodate some degree of objectivity. I suggest that we do without the notion of entitlement-preserving (but not commitment-preserving) inferences, at least outside the

domain of practical reason, and focus on commitment-preserving inferences⁴² – bearing in mind that some of them will be non-monotonic.

In the next sections I will discuss two problematic points for inferentialist semantics, on the one hand, its holistic nature, and on the other hand, its apparent difficulty in account for objectivity.

3.6 Inferentialism and semantic holism

According to inferentialist semantics, the content of some claim is determined by what claims it follows from, and what further claims follow from it. However, the inferential consequences of a given claim depend on what other claims are available as collateral premises. Therefore, the content of a claim would seem to depend on the collateral premises available: the introduction of a new premise has the potential of changing the inferential role of all other claims. For instance, the claim “The clothes are hanging outside” does not entail, on its own, “The clothes will get wet”, but this conclusion is entailed after the introduction of the collateral premise “It is raining”.

So, it would seem that whenever an agent becomes entitled to accepting some new claim, the inferential role of all the other claims accepted by that agent will change – and, as a result, the *content* of those claims would also change. This is a bad consequence for inferentialism, as Fodor and Lepore have repeatedly argued (2001, 2007), because it would rule out the possibility of shared content among agents who do not accept the exact same claims. If you do not know that it is raining, but I do, you and I would assign different contents to the claim “The clothes are drying outside”.

⁴² One may argue that, in general, inferences are not commitment-preserving, since surely agents cannot be obliged to draw all the consequences of each claim they accept (there will be an indefinitely large of such consequences). However, note that I am not saying that agents ought to explicitly endorse all these conclusions, or have active beliefs about them. It is enough if they are implicitly committed to those conclusions, so that they accept them, for instance, when questioned about them. Anyway, the same worry would apply to deductive reasoning, which Brandom regards as commitment-preserving.

Fortunately, there is an easy solution. One can think of the content of a claim as a function taking as argument the collateral premises available and yielding the inferential significance of the claim (i.e., the claims that follow from it, and those that entail it). Knowing the content of some claim would amount to knowing what inferential links it is involved in, *given the collateral premises available*. Brandom suggest this possibility – he calls these functions 'intensions' –, but seems to eventually reject it (1984: 182-184). His main reason for doing so is that it imposes a too demanding constraint over communication. According to Brandom, it should be possible that two agents express the same concepts in their speech even if they do not associate exactly the same intensions to the claims involving those concepts. For instance, physicists one hundred years ago would endorse many inferences about electrons which are rejected nowadays – however, it seems unnatural to say they were not talking about electrons.

I think that Brandom's worry can be overcome – actually, in order to do so it is enough to turn to some of the things that Brandom himself has to say about communication. Brandom argues that, in order to grasp a concept (or a claim), an agent does not need to show a perfect mastery of its norms of use: what is required is that the agent makes enough 'right moves', so that she counts as being *subject to those norms of use* (Brandom, 1994, 636). In general, it is a flexible matter how many – and which – right moves you need to make in order to be regarded as part of the game of using a given concept. In this way, Brandom allows for local deviances. It is possible that two agents have different views on the goodness of some of the inferential links involving a given claim or concept, and nonetheless they are still regarded as using the same claim or concept. In other words, it is possible to disagree about the correct use of a given claim without stopping to be dealing with a common claim – as long as there is enough agreement underlying such a discrepancy.

With this account of communication in mind, there seems to be no problem in thinking of the semantic content of claims as an 'intension' – a function from collateral premises to inferential significance. Users of a claim do not need to have a perfect grip on this function, but only a good enough one.⁴³

⁴³ This approach can be also used to answer a related objection to inferentialist theories, posed by Timothy Williamson (2003). Williamson argues that there is *not any particular inference* whose acceptance is

There are further consequences of semantic holism that should be explored. For instance, one may wonder how it affects compositionality (Fodor and Lepore, as expected, think that badly), and whether this creates problems for inferentialism. However, as I will not be examining inferentialist sub-sentential semantics,⁴⁴ I will not discuss this issue here. Instead, I will end this chapter by addressing a question which has a more direct bearing on my overall project: how inferentialism may incorporate objectivity, so as to provide a satisfactory model of our practices of describing and representing the world.

3.7 Inferentialism and objectivity

On the understanding of the notion of objectivity that I will focus on here, some state or performance may be said to be bound by an objective standard of correctness only if its correctness does not depend (or not only) on the attitudes of the participants of a relevant practice, but on the way the world is. It seems that representational practices – practices in which speakers describe how things are in the world – require this kind of

necessary for the possession of a concept. The inferentialist may reply that a speaker will count as possessing some concept insofar as there is a large enough overlap between the inferences involving the concept accepted by her and by the other speakers – but that there does not need to be a perfect match. It would not be surprising if, in many cases, there is not a particular inference that is accepted by *all* the competent users of a given concept

⁴⁴ For Brandom's treatment of sub-sentential semantics, see Brandom (1994: ch. 6); Horwich offers an alternative meaning-as-use account of sub-sentential semantics (Horwich 2005).

In section 3.5.1 I argued that inferentialists in a position to accept the assignments of truth-conditions provided by formal referentialist semantics. Thus, inferentialism can capture the sort of compositionality associated with these assignments. Of course, in an inferentialist approach the content of claims may also be determined by further entailments that are not fully amenable to compositional analysis (at least, not in interesting ways). Take, for instance, the entailment between "It is raining" and "Roads will be slippery". To the extent that learning this sort of inferential entailments is not a compositional capability, and that these entailments contribute to determining content, mastering a language will not be only a matter of grasping compositional rules. Remember, at any rate, that inferentialism does not demand that competent speakers are able to actually endorse all good entailments characterizing the content of the claims they use. Furthermore, any theory has to explain how we are able learn these inferences that do not rely on compositional capabilities.

objectivity, at least in standard cases of representation of the world.⁴⁵ If some agents are to be counted as talking about the way the world is, then the correctness of some of their performances – say, their assertions – should be determined by the way things actually are – and not merely by the opinions or attitudes of the agents engaged in the practice. Otherwise, the very notion of representation vanishes (and I will take as uncontroversial that humans do engage in representational practices).

Of course, this does not mean that one must endorse representationalist meta-semantics. An alternative is to reproduce and explain these representational practices within an inferentialist framework. On the face of it, inferentialism does not need to have especial difficulties with objectivity. One could just stipulate that the correctness of the moves of discursive practices is not determined by what agents think is correct (so, for instance, once it is decided that some moves count as expressing different mathematical concepts, the objectivist could say that it is not up to the participants of the practice whether a certain instance of mathematical reasoning is correct or incorrect).

However, as pointed out in section 2.5, problems arise when one adopts what I called pragmatism about norms – that is, a view in which being correct is explained in terms of being *treated as correct* by the participants of the practice. This pragmatist approach certainly illuminates the notion of correctness, and its place in the practice, but it runs the risk of making objectivity unreachable. In different places, Brandom argues that the norms operating in the practice should be seen as arising from the evaluations and attitudes of agents, an idea synthesized in his claim that normative statuses are

⁴⁵ That is, the sort of representation involved in descriptions of the number of chairs and tables in my kitchen. Perhaps there are special sorts of representation which lack objectivity. A possible example could be found in talk about taste, or in moral discourse. One might argue that in these domains speakers represent the world – they attribute moral properties, or the property of being tasty, to things –, but that the correctness of such representations is determined by the attitudes of those speakers – so, in these views, the attitudes of speakers would determine whether something actually has a certain moral property, or whether something is actually tasty. I do not want to commit myself here to the representational character of this kind of non-objective discourse (for relativist accounts of several kinds of assertive discourse, see MacFarlane 2014). I only need to claim that some central kinds of representation require objectivity.

instituted by the attitudes of agents (1994: 133-134, 161). This view clearly dominates the first half of *Making It Explicit*.

But Brandom is also committed to preserving a strong notion of objectivity, which allows for certain performances to be correct, or incorrect, even if none of the participants of the practice take them to be so. There are moves in the practice that answer to the way the world is – rather than the way the world is *taken to be*. Thus, Brandom insists in talking about normative statuses that may outrun the actual attitudes of every speaker in the community. For instance, an agent may undertake some commitment without acknowledging it, just because it follows from her other commitments – and this may be so even if nobody else attributes such a commitment to her. Objective content is possible because normative statuses can ‘transcend’ the attitudes that institute those statuses in the first place (Brandom 1994: 63, 137, 631). So, according to Brandom, contents would ‘codify proprieties of scorekeeping’, rather than what agents *actually do* when keeping score of each other; Brandom stresses this point by saying that his phenomenalism about norms is, after all, a ‘normative phenomenalism’ (1994: 627).

The combination of the first versions of pragmatism (phenomenalism) about norms with the later normative phenomenalism is one of the most controversial aspects of Brandom’s project and it has been widely discussed.⁴⁶ Clearly, there is a tension between these two strands. Some authors (Rödl 2010; Hattiangadi 2003; Glüer and Wikforss 2008) think that this tension reveals a contradiction: the two positions cannot cohabit. If normative statuses are instituted by attitudes, then statuses cannot override such attitudes. Others try to find a more positive getaway. Grönert (2008), for instance, suggest that Brandom’s primitive phenomenalist practices do not aim at offering a complete reconstruction of fully developed languages with objective conceptual contents. Rather, they are mere ‘paradigms’, models that help us to progress but that have to be left behind by the time we arrive at mature discursive practices. This is why phenomenalism about norms ends up by transmuting into normative phenomenalism⁴⁷

⁴⁶ Shapiro (2004); Wanderer (2008: 78-92), Laurier (2005), Grönert (2005), Rödl (2010), Rosen (1997).

⁴⁷ A similar line is pursued in Wanderer (2008: 89-93). Here, the normative game model is just a ‘stalking horse’ that, instead of offering a reductive explanation, helps us to explicate our linguistic practices (which

(note that even in these sympathetic reading the two kinds of phenomenalism are judged to be incompatible).

I think it is safe to say that Brandom's views on this point require further development, in order to form a consistent whole. This is more so if one takes seriously the pragmatist constraint that normative statuses must be explained in terms of their practical significance – in terms of their role in the activities of agents. As it stands, it is not clear how Brandom's proposal would manage to fulfill its promise of getting norms to be domesticated or made non mysterious – it seems that, ultimately, statuses would be understood in terms of *appropriate* attitudes, so in the end the situation is that statuses are explained by further statuses (Brandom 1994: 626-627; Rödl 2010: 75-76).

Brandom acknowledges these difficulties, but he still thinks that it is possible to construct discursive practices giving rise to objective norms, starting just with the materials offered by a 'phenomenalist' or pragmatist analysis of the institution of norms – that is, an analysis of norms as established by the attitudes of agents (Brandom 1994: 54, 155-156). Let me sketch briefly Brandom's proposed solution, in order to assess its limitations. In Brandom's view, objectivity imposes the following constraint: there should always be room for the distinction between what a given agent (or group of agents) takes to be correct and what is actually correct (Brandom 1994: 592-607). So, nothing is right *just in virtue of being regarded as right from some perspective*: every perspective is in principle fallible (Brandom 1994: 592-607).

Brandom argues that this objectivist constraint can only be satisfied in a *social* practice, in which there are multiple agents assessing each other – that is, multiple assessing perspectives. A social practice with a structure allowing for objectivity will be such that what is regarded as correct from the perspective of a given agent may always be regarded from the perspective of a second agent as merely *taken to be correct* by that first agent. In this way, there is always, in principle, a distinction between what is actually correct and what is merely taken to be correct by a particular agent. No perspective – not even the perspective of the whole community – is privileged, in the sense of making some view correct just because it is taken to be so from that perspective (Brandom 1994: ch. 8).

means: helps us to express explicitly what is involved in a practical mastery of our language). Wanderer admits that this is not Brandom's official position.

It may very well be that this sort of perspectival articulation is a *necessary* condition for objectivity, but it seems clear that it does not amount to a *sufficient* one. In Brandom's picture, one finds a multiplicity of perspectives assessing each other, so that the claims made from some perspective may always be seen as wrong from a different perspective; however, there is no reason at all to think that these practices will give rise to any kind of objective correctness. It is not clear how this swarm of potentially clashing perspectives could converge on anything that deserves to be identified as what is actually the case. In particular, Brandom has not shown how the world acquires authority over practices with this kind of structure – how the correctness of some of the moves of these practices would be determined by the way the world is (see Prien 2010 for this kind of objection).

This shortcoming in Brandom's proposal can be made more vivid by noting that the sort of social structure which Brandom associates with objectivity could be taken to be deployed by a discursive practice with relativist semantics in which truth were relative to the assessing attitudes of evaluators (I will develop this point in chapter 8). Assume that discourse about taste has relativist semantics – as constructed by MacFarlane (2014). More specifically, two different agents asserting “Apples are tasty” would be expressing the same content, but the truth (i.e. correctness) of each assertion of that claim would depend on the standard of taste operative in the relevant context of assessment (see MacFarlane 2014);⁴⁸ moreover, assume that the standard operative in the context of assessment is determined by the gustatory attitudes of the agent *evaluating* the assertion. So, Jane's assertion “Apples are tasty” may be true in relation to Martin's perspective (he likes apples) but false in relation to Jane's (she finds them dull). And, of course, there is no fact of the matter as to which of them is *objectively* right: whether Jane's assertion is correct, as evaluated from a given context of assessment, depends on the gustatory attitudes of the evaluator (and remember that I have claimed that objectivity requires that correctness is not determined by the attitudes of evaluators).⁴⁹

⁴⁸ According to moderate versions of relativism, the correctness of an assertion must be evaluated with respect to the parameters (e.g. taste standards) relevant from the context of utterance – rather than the context of assessment (see Kölbel 2015a,b).

⁴⁹ Note that this does not mean that the notion of objectivity has no room in any relativist discursive practice. There may be relativist practices in which truth is made relative to some feature of the context of assessment that is not determined by the attitudes of the evaluators occupying such a context: for instance,

Relativist practices like this, however, seem to meet Brandom's objectivistic requirements, since what is correct from Jane's point of view (i.e. in relation to her gustatory attitudes) may always be seen as a mere incorrect opinion from a different perspective – say, Martin's. Nothing is infallibly right just in virtue of being regarded as such from a specific perspective. Thus, Brandom's constraints on the structure of social practices do not manage to characterize what is for an agent to take some claim as being *objectively* right or wrong – since they are satisfied by this sort of non-objective relativist speech (I will discuss this issue in more detail in section 8.4).

Brandom argues that, in order to find out what is actually the case, agents will have to engage in 'a messy retail business of assessing the comparative authority of competing evidential and inferential claims' (1994: 601). However, in *Making It Explicit* (and *Articulating Reasons*), he does not explain in virtue of what this process of investigating and assessing alternative views is sensitive at all to the way things are in the world. Such an explanation is sketched in very vague terms in more recent work, particularly in *Between Saying and Doing* (2008). There, it is argued that agents update their views by engaging in perception-action feedback loops (I will discuss Brandom's view's in more detail in section 4.5). I think this proposal is certainly promising, but it is difficult to know what to make of it in the embryonic state in which Brandom presents it. In particular, it would be necessary to examine whether a proposal of this sort may overcome the kind of difficulties faced by similar pragmatist approaches, including difficulties that have been highlighted by Brandom himself – for instance, in his criticism of success semantics (Brandom, 1994b; I will discuss these objections in chapter 5).

the time of that context, or the body of evidence accessible. In this sort of relativist practices, it may be an objective question whether some assertion is correct as assessed from a given context: that is to say, whether the assertion is correct, as assessed from that context, may not depend on the assessing attitudes of the evaluators occupying the context. I am only arguing that one can construct relativist practices in which truth is made relative to the attitudes of evaluators (say, their gustatory attitudes), and that these practices seem to meet Brandom's criteria for objectivity– despite being clearly non-objective, at least in the sense of objectivity relevant for the present discussion. Perhaps these relativist practices are not a good model of any real human discursive practice, but they are at least possible, coherent practices and that is enough to put Brandom in trouble.

My main aim in the following chapters will be to flesh out this kind of pragmatist approach to representation. I will try to develop an inferentialist account of representational practices which manages to incorporate objectivity – at least, the sort of objectivity that characterizes human language and cognition – while refraining from introducing notions that are not explained in terms of what agents do in their practices – so that the project remains loyal to its animating pragmatist spirit.

3.8 Conclusions

In this chapter I have argued that inferentialism offers the best framework for developing a practice-based account of meaning. Focusing on Brandom's version of inferentialism, I have explained in some detail how inferentialist semantics can be seen as grounded in what the participants of discursive practices do, thereby satisfying a practice-based meta-semantics. I have also reviewed some difficulties faced by this approach, and have indicated two points in which Brandom's proposal is unsatisfactory as an account of representational practices. First, his notion of entitlement is too weak, since it allows for proper entitlement to false descriptions of the world. Second, Brandom's social construction of objectivity does not explain how the world acquires authority over social practices. In the following chapters, I will try to develop an inferentialist, practice-based, account of representational practices. In order to do so, I will explore in which way discursive practices involve interactions with the world, and how the outcomes of these interactions may have a saying in determining the correctness of the moves of such practices. As a first step, in the next chapter, I will examine an inferentialist approach to practical reasoning and observation.

Chapter 4: Perception and Action in Discursive Practices

4.1 Introduction

On the view I am proposing, discursive practices can be said to represent the world (so that the correctness of moves within the practice is determined by the way the world is) by virtue of involving interactions with the environment – interactions whose outcomes depend, at least partly, on how things stand in the world. These interactions with the environment include observation and action (in particular, non-linguistic action, that is, the performance of actions which are not speech acts). Thus, a practice-based account of representation must examine the role of observation and non-linguistic action in discursive practices.

The aim of this chapter is to lay out an account of how observational and practical dealings with the world can be integrated within the sort of inferentialist model of discursive practices introduced in the previous chapter. As I already mentioned there, Brandom (1994: ch. 4; 2000) allows that observation and action play an important role in discursive games of giving and asking for reasons. More specifically, according to Brandom, one may acquire reasons to accept certain claims (e.g. that there is a red table in front of you) by virtue of being in certain observational circumstances (e.g. seeing a red table in front of you); likewise, having reasons for accepting certain claims (e.g. that it is raining) may give your reasons for performing certain actions (to open the umbrella). Brandom argues that these connections with observation and action are part of what makes claims have the content they have.⁵⁰

In the first part of the chapter I discuss in more detail how observation and action are related to discursive moves – in particular to the endorsement of claims, as manifested via assertions. I discuss observation in section 4.2 and action in section 4.3. I will take Brandom's model as my starting point, although I will depart from it in several

⁵⁰ Although he allows for the possibility that claims in certain domains (e.g. pure mathematics) acquire their contents entirely by virtue of their inferential connections with other claims (Brandom 1994: 167). The relations to perception and action are essential, however, for characterizing claims with empirical content.

points. In the second part of the chapter I examine how the outcomes of our practical interactions with the world may be evaluated, within an inferentialist framework. On the view I will present, actions are assessed as failures when one observes outcomes that are incompatible with the success of the relevant actions. In section 4.4 I argue that, when an action fails in achieving its goal, it is because the agent actually lacked proper entitlement for the performance of the action (as a means of achieving her goal). In section 4.5, I critically assess the role of the notion of practical success in Brandom's theory. Drawing on the discussions on these two sections, I develop in section 4.6 a model of action-evaluation feedback loops: in such loops, agents evaluate the outcomes of their actions and update their beliefs accordingly.

4.2 Perceptual entitlement

In the previous chapter, I presented an inferentialist model of discursive practices (following Brandom) according to which claims are characterized by their role in practices of giving and asking for reasons – more specifically, by the commitments brought by an endorsement of such claims and by the ways in which one could obtain entitlement to those commitments. In this model, the paradigmatic way of vindicating one's entitlement to endorsing a given claim is by showing one's entitlement to endorsing some further claim related to the first one through suitable inferential connections (e.g. entitlement to endorsing the claim "The table is red" provides entitlement to endorsing the claim "the table is coloured").

However, in some cases entitlement to endorsing a claim is not earned by virtue of being entitled to endorsing some other claim entailing it, but rather because one's endorsement of the claim is elicited by the operation of reliable perceptual capacities. The reliable functioning of our perceptual capacities entitles us, in favourable circumstances, to accepting certain claims (Brandom 1994: ch. 4). Brandom calls this sort of claim 'observation reports' and, drawing from Sellars (1954), regards them as resulting from entry moves in the inferential games of giving and asking for reasons (Brandom 1994: 213-229). These reports belong to the game of giving and asking for reasons, insofar as they can be used to provide reasons for the performance of further moves;

however, they result from entry moves in the sense that entitlement to making them is not provided by other moves in the game (i.e. other entitled assertions), but by the non-inferential exercise of suitable perceptual sensibilities.

So, for instance, seeing that the cat is in the garden – you are standing in front of it, in good light conditions – will generally entitle you to accepting the claim “The cat is in the garden”. The responses of your perceptual sensibilities – which, from where you stand and in those light conditions, are reliable in discriminating situations in which the cat is in the garden – make you count as entitled to endorsing the relevant observation report (i.e. “The cat is in the garden”). I will call entitlement acquired in this way *perceptual entitlement*.

This does not mean that observation reports are infallible or that our perceptual capacities have some form of unquestionable normative authority (Sellars (1956) insists on this point). Notice that I have been saying that the responses of our perceptual capacities only bring entitlement to making observation reports insofar as such capacities are *reliable* and are working in *favourable conditions*. So, any particular observation report may be challenged if it happens to be elicited by the exercise of unreliable perceptual capacities or if environmental conditions are not favourable). One may only attribute perceptual entitlement to an agent if one takes her perceptual capacities to be reliable in those conditions of operation (see Brandom (1994: 206-227) for this view).⁵¹

⁵¹ Sellars (1956) argues that, in order for an observation report to be entitled, it is the agent exercising the relevant perceptual capacities who has to acknowledge that such capacities are reliable. Brandom (1994: 215-221) is less demanding: he allows for situations in which a report is entitled even if the attribution of reliability to the relevant perceptual capacities is not made by the observer exercising them, but by the scorekeeper attributing the entitlement to the report.

I will not discuss this issue in detail here. Let me just make some brief points. Perhaps there exist reasons for an agent to endorse a certain observation report as long as the relevant perceptual capacities are actually working reliably (and regardless of whether the agent exercising the capacities acknowledges herself such reliability); but, unless the agent making the report regards the relevant capacities as reliable, she will not be in a position to *possess* such reasons to accept the report (for the difference between there being reasons/warrants for something and an agent possessing such reasons/warrants, see Whiting 2013).

Note that the fact that a particular perceptual response is veridical does not mean that the operations of such perceptual capacities are reliable: reliability requires that the relevant perceptual capacities tend to produce the correct output, so that such correct outputs are not just the result of luck. So, even if a certain perceptual response is accurate, if the operations of the perceptual capacities are not reliable in those conditions (e.g. the response was accurate merely out of luck) it can still be that such a response does not provide the observer with entitlement to endorsing the corresponding observation report.

A stringent way of constructing the notion of reliability is to demand that reliability implies perceptual success (that is, observations that are veridical). According to this understanding of the notion of reliability, whenever some perceptual capacities fail to detect accurately environmental features, they will be taken to be working unreliably. On this view, the responses of reliable perceptual capacities will always provide proper entitlement to endorsing the corresponding observation report.

On a less stringent understanding, perceptual reliability does not imply perceptual success: reliable perceptual responses may be inaccurate (e.g. if one is unlucky or the environment happens not to collaborate). On this view, perceptual reliability does not need to provide (proper) entitlement to endorsing the observation reports elicited by the operations of the relevant perceptual capacities. Proper entitlement to endorsing such observation reports requires not only the reliability of the relevant perceptual capacities, but also their accuracy (the report has to be true). Thus, the inference from “That observation report has been elicited by the responses of reliable perceptual capacities” to “The speaker was entitled to making that observation report” is defeasible. There can be circumstances that defeat the link between perceptual reliability and perceptual entitlement.

For my purposes here, one can choose either of these two understandings of the notion of reliability. What I will require is that perceptual entitlement implies perceptual success (that is, perceptions that are accurate). In this way, an agent properly entitled to making some observation report cannot be making a false or inaccurate claim. I depart in this point from Brandom, who seems to allow for the possibility of proper entitlement to false claims (1995; 2000: 201). Roughly, my reason for rejecting this possibility is that for a

discursive practice to be representational, false claims must count as incorrect in some way. Given the normative materials I am building my account with, a claim can only count as incorrect if there is no entitlement to making it (I discuss this issue in more detail in section 8.3.3).

A possible worry is that it may seem that one will never be in a position to appeal to perceptual entitlement in order to vindicate one's entitlement to making a certain claim. The possibility of some ignored impediment (say, hallucinatory drugs or deceiving mirrors) getting in the way and thwarting the operations of our perceptual capacities can never be fully discarded. However, this is not a problem if one regards attributions of perceptual entitlement as *defeasible* (Williamson 2001). In general, one may attribute perceptual entitlement if there are good enough grounds for doing so and one has no reasons for thinking that any obstacle or unfavourable circumstances have been hindering the proper functioning of the relevant perceptual capacities. If, after all, it turns out that some such defeating conditions were actually taking place, the attribution of perceptual entitlement will be revoked. But this does not mean that the attribution of perceptual entitlement is inappropriate in cases where no defeating condition takes place (even if one has not discarded explicitly all possible defeating conditions).⁵²

In particular, the functioning of the perceptual capacities of typical agents tends to be regarded as reliable in standard conditions – so, if someone looks out the window and says that it is raining, we will typically think that her perception is reliable and the corresponding report is entitled. Brandom (1994: 176-178; also Williams 2001) argues that attributions of entitlement in general have a 'default and challenge' structure. According to Brandom's model, many of the claims made by agents are attributed entitlement by default, unless there are reasons for doubting them. When such reasons for doubting exists, the agent's views may be properly challenged (or at least questioned), and she will have to vindicate her entitlement to such views in a suitable way. Perceptual entitlement seems to work in accordance to this default and challenge model (as Brandom (1994: 222) points out). By default, we assume that the responses of the perceptual sensibilities of normal individuals, working in ordinary conditions, are reliable

⁵² The defeasibility of attributions of entitlement in human representational practices will be important when discussing the possibility of objectivity in chapters 7 and 8.

and accurate. Accordingly, by default we attribute entitlement to observation reports made in ordinary conditions by normal individuals. The reliability of the responses of the perceptual capacities of normal individuals will, in general, only be questioned when there are reasons to think that some obstacle for the proper functioning of those capacities is taking place (say, light conditions are deficient or we know that the agent is under the effects of some drug). Thus, Brandom's default and challenge model provides an account of how defeasible attributions of perceptual entitlement may work.

Attributions of perceptual reliability do not need to be made explicit (Brandom 1994: 218). Usually, such attributions remain implicit in our assessment of the agent's observation report as entitled. By attributing entitlement to your observation report that it is raining outside, I am implicitly treating the operations of your perceptual system – which underlie such a report – as reliable. Again, it is only when there are reasons not to take for granted the agent's perceptual reliability that such reliability will need to be explicitly considered. So, if you claim to be seeing that the cat is in the garden, but it is a moonless night and the lights are off, someone may properly challenge the reliability of your vision in those conditions – and you may try to vindicate such reliability, for instance by pointing out that you wear night-vision goggles.

To sum up: following Brandom, I have argued that observation may be integrated into an inferentialist model of discursive practices by treating observation reports as resulting from entry moves into such practices. Entitlement to making these reports is not gained inferentially – by appealing to one's entitlement to some other discursive moves –, but rather by virtue of such reports being elicited by the operations of reliable perceptual capacities. Moreover, I have argued that perceptual entitlement is infallible – in the sense that one cannot be perceptually entitled to endorsing a false claim. However, *attributions* of perceptual entitlement are not infallible, but defeasible and subject to revision.

In the next section I will discuss non-linguistic action as an exit move in discursive practices.

4.3 Intentional agency and practical reasoning

4.3.1 Intentional agency

According to Brandom (1994: 229-271; 2000), intentional agency involves the undertaking of practical commitments – commitments to the performance of some action. A commitment to performing a given action (say, to opening the door) is satisfied by performing such an action (by closing the door).

On Brandom's view, the acknowledgement of a practical commitment can take two shapes (Brandom 1994: 256 - 259): on the one hand, the performance of an action can count itself as the acknowledgement of the practical commitment that would be fulfilled by that action. My opening the door can be taken as an acknowledgement of the commitment to opening the door. This would be what Searle called 'intention in action' (Searle 1983). On the other hand, I can acknowledge a commitment to perform some action when certain conditions occur (without performing the action at the same time that I acknowledge the commitment). This would be a 'prior intention'. For instance, if at 7 a.m. I say 'I shall take the 9 a.m. train',⁵³ I am expressing my intention to do something in the future, and I thereby commit to performing such an action when the right time comes. A prior intention, thus, is fulfilled by a further intention in action. Note, however, that sometimes the absence of a suitable intention in action does not entail that we are behaving improperly: maybe the satisfaction conditions for the previous prior intention are never met, so we are not in a position to fulfil our commitment. Imagine that I say 'I shall give a special present to the first customer to enter the shop today', but then no customer enters the shop today.

When undertaking a practical commitment, it may be questioned whether one is entitled to such an undertaking. Intentional agency, in this way, is responsive to reasons: agents acting intentionally may be asked for their reasons to do what they are doing. This is what Anscombe argues when she claims that intentional actions answer to (justificatory) why questions (Anscombe, 1957). So, while intentional action involves the acknowledgement of a practical commitment, not every practical commitment will be

⁵³ I am following Brandom's convention of expressing the acknowledgment of practical commitments with the locution 'I shall...!'.

entitled: we can act intentionally without good reasons (or even without reasons at all). What makes an action intentional is not that there are reasons supporting it, but that such reasons may be demanded.

4.3.2 Practical reasons, facts and attitudes

How is practical agency connected with inferential discursive practices? In Brandom's view (1994: 243-259; 2000), entitlement to undertaking discursive commitments (i.e. to endorsing certain claims) may provide entitlement to undertaking practical commitments. Thus, according to Brandom, one may vindicate one's entitlement to the performance of an action by appealing to one's entitlement to accepting certain claims – that is, by making certain claims to which one is entitled. In order to vindicate my entitlement to turning the lights off I may point to the fact that nobody is in the room (or, in a more roundabout way, to my entitlement to treat the claim that nobody is in the room as a fact).

This inferentialist model of the vindication of practical entitlement may be integrated within a general account of practical reasoning, according to which normative (i.e. justificatory) reasons for action are facts supporting or recommending the relevant actions (Álvarez 2010; Vogler 2001; Dancy 2000, 2004). Within this approach, in order to give reasons for one's actions, one would cite facts that support those actions. In this way, entitlement to accepting certain claims as facts may be seen as providing entitlement to performing certain actions – namely, the actions for which those facts are reasons –, insofar as proper entitlement to accepting a claim implies that such a claim is, or expresses, a fact.⁵⁴ The crucial point is that the reasons for acting are not constituted by the agent's endorsing some claim, but rather by the claim endorsed (or, if you prefer, the fact such a claim constitutes or expresses). So, in the approach I will follow, having a given belief (believing that it is raining) is not a normative reason for performing some action (opening the umbrella): the reason for acting is the fact believed (that it is raining).

⁵⁴ Remember that I am holding the view that one can only be properly entitled to accepting accurate claims – claims that accord with the way things are (see chapter 2; and chapter 7 and 8 for further discussion).

For the sake of clarity, let me specify how I will understand the distinction between normative and motivating reasons. I will think of a normative reason as a reason that justifies the performances of an action (i.e. makes such a performance entitled). On the view I will endorse, a normative reason will also be motivating when the agent acts in light of such a reason – that is, when the agent takes it as her reason for acting. In order for a reason to be motivating, it may be necessary that the agent has certain attitudes (e.g. certain beliefs), but I will take it that the reason is constituted by the considerations⁵⁵ towards which those attitudes are directed, and not by the attitudes themselves (see Álvarez 2010).

The approach I have sketched runs against the view that mental states, and in particular desires and other pro-attitudes, provide normative reasons for action⁵⁶ (Gibbons 2010; Williams 1985; for criticism of this view, see Dancy 2000; Parfit 2011). On such a view, having a desire to eat an apple, and having the belief that there are apples in the kitchen, would be reasons recommending my going to the kitchen and take an apple (i.e. bringing entitlement to my performing such an action). My desire to eat the apple would be part, on this view, of what brings entitlement to my action of going to the kitchen – it might even be thought that there will no reasons recommending an action in the absence of desires or some suitable pro-attitude.

As I said, I will endorse an alternative approach, in which only facts are (normative) reasons. In this way, Álvarez (2010) argues – following Anscombe – that desires themselves never act as normative reasons for our actions: when we desire something, we usually desire it for some reason, because there is some feature in the desired thing that makes it desirable, that makes it good for us. It is these 'good-making' aspects of what we desire that recommends (entitles) our actions – and when vindicating our entitlement to performing an action, we appeal to such good-making considerations. So, a piece of practical reasoning would go like this: "I am hungry, eating a sandwich will satisfy my hunger; crisps are cheaper, but they will only make me hungrier. Therefore, I

⁵⁵ Although I am sympathetic to the view that only factual considerations can be real motivating reasons, I will allow for the possibility that false claims may also be motivating reasons. At any rate, motivating reasons will not be mental attitudes, but claims or considerations.

⁵⁶ 'Desiring something is of course a reason for doing it' (Williams 1985: 19).

shall eat a sandwich”, or “If I lend the book to my colleague, I will be helping her, so I shall lend her the book”. In these cases, satisfying hunger and helping a colleague are regarded as something good (something worth aiming at), and as a result the agent takes it that there are normative reasons for performing actions with such good outcomes: the agent sees herself as entitled to performing the action because of these reasons.⁵⁷ However, desires for good things are never normative reasons themselves – and *the fact that we desire* those good things does not typically appear among our reasons for acting.⁵⁸

One clear motivation for this approach is that, arguably, desiring something that is not actually desirable will not give you good reasons to act.⁵⁹ So, imagine that, out of spite, Mary desires to harm her colleague James, and therefore destroys his very important manuscript. Looking back, she realizes that it was not right to desire to harm James; thus, she will now think that she did not actually have good reasons for destroying his manuscript. Actually, there were good reasons not to destroy James’s manuscript, even if Mary lacked the desire not to destroy it. Otherwise, just by having the desire to torture innocent people you would have good reasons to harm innocents – and you could claim that there are no good reasons for you to help someone in need, just by arguing that you do not have the relevant desire.⁶⁰

⁵⁷ These reasons are normative (insofar as they provide entitlement to endorsing the conclusions), even if also motivating (the agent takes them as her reasons for acting).

⁵⁸ The *fact* that I have some desire can constitute a reason for action in some special cases. For instance, the fact that I always desire to procrastinate, which is bad for my work, may give me reasons to go to work to the library (where there are fewer opportunities for procrastination). See Álvarez (2010).

⁵⁹ Notice that this does not need to prejudice any debate about objectivism or realism concerning values or moral obligations. It could be that the desirability of some end (as pursued by a given subject) were relative to the perspective from which such desirability is assessed. This would require accepting a relativist view of value facts and truths, but such a relativist view is an open, even if controversial, possibility that finds advocates in the literature (see López de Sa 2010). One could further argue that whether something is desirable from a certain perspective is determined by the desires (or in general pro-attitudes) *actually* had by the individual occupying such a perspective – or perhaps the pro-attitudes she is disposed to have when she is well-informed and has reflected sufficiently about the relevant matter.

I will remain neutral here about this sort of debates.

⁶⁰ Of course, the fact that some virtuous (or morally reliable) agent has a desire can provide good evidence about the desirability of such a desire – in the same way that the fact that a reliable observer perceives something offers reasons to regard such a perception as accurate.

Moreover, an approach of this kind – in which reasons are given by facts – seems to be required by my project of accounting for representation by appeal to practical reason. Ultimately, I will argue that, if the performance of an action is properly entitled, then such an action cannot fail in achieving its goals. This claim blends naturally with the view that the reasons providing proper entitlement for actions are facts, rather than mental attitudes. Having the belief that the tea is not very hot does not imply that I will not get burnt if I drink it: if my belief is false, I will actually get burnt. What makes it the case that I may drink the tea without burning is the fact that it is not very hot. It is this fact that provides me with reasons for drinking the tea, rather than having to wait for it to cool down.

One possible alternative would be to argue that having beliefs may provide reasons for action, but that only having *true* beliefs gives *good* reasons for action (i.e. reasons making the action properly entitled). However, I will adopt the view that reasons are facts, because of the sort of considerations put forward above (i.e. when justifying our actions we do not appeal to our mental states, but to facts, and we often acknowledge that there were actually reasons for having done something in the past even when one did not have the relevant beliefs in that past moment).

4.3.3 The expressive role of attributions of pro-attitudes and goals

It is common, within the sort of Anscombian approach I am presenting, to argue that goals or ends are not reasons and do not figure among the premises of practical reasoning. So, for instance, Vogler (2001: 458) writes that 'Anscombe has argued that mention of wants or ends has no place among the premises of a practical inference' (2001: 458). Likewise, according to Álvarez (2010) goals are the 'beginning' of our practical reasoning, they tell us about its 'point'; as such, it will generally be useful to present them in addition to the premises of the inference – but not *among them*. Specifying our practical goals will help to determine when to stop the reasoning, and 'how things are supposed to come out' according to our reasoning (Vogler 2001: 458).

Brandom (1994, 2000) presents an alternative proposal. He argues that specifications of pro-attitudes and goals have an expressive function – similar to that of

conditionals. According to Brandom, locutions specifying goals and pro-attitudes make it possible to express explicitly one's endorsement of patterns of practical reasoning – that is, of pieces of reasoning connecting certain premises with certain practical conclusions. For example, if I express the desire to stay dry, I will become committed to regarding the fact that it is raining, and that only opening my umbrella will keep me dry, as giving me good reasons for opening my umbrella. My expression of a desire to stay dry will also commit me to endorsing other related pieces of practical reasoning, such as: "I will only keep dry if I take the bus; then, I will take the bus." Each goal, thus, underwrites the endorsement of a whole family of patterns of practical reasoning – more specifically, those patterns of reasoning that recommend courses of action that lead to the fulfilment of the goal.

So, locutions specifying goals, according to Brandom, allow us to make explicit features of the inferential articulation of our practices of giving and asking for reasons. Consequently, these locutions are regarded by Brandom (1994: 247-249) as a form of logical vocabulary – remember that, in Brandom's view of logic, this expressive function is what characterizes logic in a broad sense. Thus, by resorting to vocabulary expressing goals and pro-attitudes one can make assertions saying explicitly what otherwise can only be done implicitly – namely, to endorse some patterns of practical reasoning. In this way, one can question explicitly whether some agent is entitled to endorsing certain patterns of practical reasoning, by questioning whether she is entitled to accepting the corresponding goal. Once someone's acceptance of a given goal is expressed in a propositional, assertable form, one can assess reasons in favour and against such an acceptance.

I think that Brandom offers an interesting and promising account of the role of specifications of goals and pro-attitudes. However, as I see it, his specific proposal has some shortcomings and can be improved. Brandom (1994: ch. 4) chooses the locution 'Let me ...' to express desires – 'Let me stay dry', therefore, would be used to express my desire to stay dry. It may very well be that such locutions express desires, but clearly they do not do it in an assertive way. 'Let me stay dry' is not a declarative sentence that can be used to make assertions. Lacking the potential to be asserted, expressions of desires would not be in a position to be involved in inferential practices of giving and asking for reasons – which was the main motivation for having locutions with such a logical-expressive role.

So, I think it is better to choose a locution that can be used to make assertions: for instance, 'It is desirable (for me) to stay dry' (or perhaps 'It is good (for me) to stay dry' or 'I should stay dry').⁶¹ These locutions can be used to make assertions about the desirability of my staying dry – assertions for which reasons can be asked and given. Furthermore, this way of expressing ends makes it clear that what is relevant for the proper entitlement of my actions is not that I happen to have a certain desire or goal, but rather that my goal, what I desired, is something actually desirable, something that actually deserves to be pursued. Compare with the claim "I have the desire to stay dry". This claim expresses the belief that I have the desire to stay dry. I will be entitled to accepting such a claim if I do actually have this desire. But the fact that I have such a desire does not need to make it something that merits being desired.

The view I am proposing, therefore, is that locutions specifying goals should have a form like 'It is desirable (for me) to...', 'It is good (for me) to...', or 'I should...'. If one is entitled to accepting the claims expressed by these locutions, then one will be entitled to endorsing the patterns of practical reasoning underlain by such goals. So, the fact that some end is desirable (for you) means that there are reasons for you to perform an action which is a necessary means for that end.⁶²

Claims about the desirability of some end purport to express facts, so they could be included among the premises of a piece of practical reasoning. However, one could still maintain – with Brandom (1994) – that the premises of a non-enthymematic piece of practical reasoning do not need to include such claims. The acceptance of the goal may remain implicit in one's endorsement of the corresponding patterns of practical reasoning. Claims stating explicitly the desirability of the goals of the agent have an expressive function, which, in Brandom's view (1994: 248), is useful but optional. In the

⁶¹ Perhaps one can judge that something is desirable without actually desiring it. But, in this case, the agent will still take it that there are reasons for desiring that thing. In the complementary case, where one desires something without finding it desirable, the agent would take it that there are actually no good reasons for desiring what she does (and, consequently, for trying to fulfill such a desire).

⁶² More specifically: when one commits to the desirability of doing x – e.g. by claiming that 'It is desirable for me to do x' or 'I should do x' – one thereby becomes committed (prima facie at least) to doing x or to performing some other action that is a good means for doing x (e.g. opening the umbrella in order to stay dry). Thus, one gets committed to endorsing pieces of practical reasoning in which some action is recommended as a good means for the desired end.

same way, according to Brandom (1994), conditional premises specifying inferential entailments do not need to figure among the premises of a piece of theoretical reasoning in order for it to count as non-enthymematic.

For my purposes here, I can remain noncommittal about the correctness of the view that non-enthymematic pieces of practical reasoning do not need to include among their premises a claim about the agent's goal. For the sake of clarity, when presenting a piece of practical reasoning I will to cite explicitly the agent's goal.

4.3.4 Objections to the expressive role of pro-attitudes

In this section I will address a possible objection to Brandom's account of the expressive (i.e. logical, in Brandom's sense) role of specifications of desires – an objection that would also apply to an attribution of such an expressive role to the sort of specifications of goals I have just presented.

The objection is the following: White (2003) claims that specifications of desires are unsuited to perform the sort of expressive role that Brandom assigns to them. The reason is that they are monadic locutions. In theoretical inferences, one's endorsement of an inferential link between claims can be expressed – in an assertable way – by means of conditionals. Conditionals are dyadic operators: they have two slots ready to be filled by the claims connected, the antecedent and the consequent. This makes conditionals especially well-suited for performing their logical-expressive role: the inferential entailment between p and q would be easily codified by means of the conditional 'If p , then q '. By contrast, the locutions used for expressing desires and goals ('Let me ...' in Brandom's view, 'It is desirable (for me) that ...' in my proposal) are usually monadic operators; because of this, White doubts that they can fulfil the logical-expressive duties Brandom asks from them: how can they relate two claims (premise and conclusion) if they have only 'one slot' available?

One first answer is that only a very formalist approach would justify the demand that a logical locution performs its expressive role in virtue of its lexico-syntactical structure. A conditional expresses an inferential connection because the norms of the practice make it so, not because of its lexico-syntactical form. Accepting "If $p \rightarrow q$ " involves undertaking an inferential commitment – a commitment to the goodness of the

inference from p to q – , and that is why that conditional plays a logical role. But another locution, say ‘The conditional x is OK’, could perform the same task (think of the variable x as naming the inferential entailment codified by the locution). This is so even if neither the expressions ‘ p ’ and ‘ q ’, nor any other expression naming p or q , are present in that alternative logical locution. When someone asserts “The conditional x is OK”, we would take her to be accepting the inferential entailment referred to by ‘ x ’, say, the entailment between p and q . Of course, the fact that a conditional is constructed by linking the two claims related by the inference makes it especially suitable for its logical-expressive function: thanks to the dyadic structure of conditionals, we will be able to encode inferential commitments in a very straightforward and systematic way. However, this is just something convenient, not necessary.

Why wouldn’t we take advantage of dyadic operators in the case of desires and goals, then? I suggest that this is because specifications of goals do not codify explicitly a single piece of practical reasoning, connecting one set of premises and some conclusions, but a whole family of patterns of practical reasoning. The goal to eat something sweet underlies a piece of practical reasoning that concludes in an endorsement of the action of eating chocolate, but also a piece of practical reasoning which licenses eating a piece of carrot cake. Dyadic operators are not especially good for this logical role – there is not a single set of premises and one conclusion, but an indefinite number of possible premises and conclusions. Expressions of goals do the best they can, which is to indicate explicitly the ultimate conditions of success of intentional actions: what kind of outcome those actions should have in order to be properly licensed by the pieces of practical reasoning underwritten by the goal. So, if I say that it is desirable for me to eat something sweet, I am endorsing *prima facie* all those pieces of practical reasoning that lead to an action as a result of which I eat something sweet.

Another way to look at it is to regard specifications of goals as playing a higher-order logical-expressive role. Whereas conditionals codify reasoning patterns connecting certain premises with a certain conclusion, a single goal would underpin several such reasoning patterns. By endorsing a certain goal, thus, one would be endorsing a collection of conditionals – the conditionals codifying all those pieces of reasoning underpinned by such a goal. The logical-expressive functions of conditionals and of specifications of goals, therefore, would take place at different levels.

4.3.5 Practical inference

Before examining how the evaluation of practical success leads to the updating of the agent's views, I will discuss in more detail the nature of practical reasoning. The main purpose of this section is to have a precise and clear understanding of the notions of practical reasoning and inference, which will play a central role in the rest of the dissertation. In particular, I will discuss whether pieces of practical reasoning are correct by virtue of some inferential relation of entailment between propositions – in the same way that the fact that p entails q makes the piece of theoretical reasoning from believing that p to believing that q good.

The question I want to address, therefore, is whether there are practical inferences. Before going any farther, one has to bear in mind that the notion of inference can be understood in different ways, as I already mentioned in chapter 3 (Harman 1984, 1986; Vogler 2001; Streumer 2010). On the one hand, 'inference' may refer to reason-responsive transitions between psychological states or attitudes (that is, episodes of reasoning). On the other hand, one may be talking instead about inferential entailments connecting propositions – i.e. logical relations or arguments. Inference, in the first sense, is something done by agents; in the second sense, it is a relation facts or propositions stand in. While it is very plausible that there are instances of practical inference in the first sense (that is, there are instances of practical reasoning), it is far less clear that there are practical inferential entailments (see Harman 1976; Streumer 2010).

The view I have been presenting is that entitlement to endorsing certain claims (certain discursive commitments) may offer entitlement to endorsing certain practical commitments. In other words, certain facts may provide reasons for endorsing certain courses of action. In virtue of this, certain episodes of practical reasoning (in which an agent goes from certain beliefs to acting, forming an intention or otherwise endorsing a course of action) count as good or correct.

Practical reasoning, as a reasons-responsive transition among attitudes (including the endorsement of actions as an attitude), is therefore not problematic in the framework I am presenting. But it is not clear whether there is an obvious place for practical inferential entailments underlying those pieces of practical reasoning. Such inferential entailments, remember, would hold between propositional contents, rather than between

attitudes toward contents. This undermines the view that desires or pro-attitudes figure among the premises of practical reasoning: desires are attitudes, rather than propositional contents (Vogler 2001). But similar worries arise in relation to the idea that there are practical inferential entailments concluding in actions or intentions – since neither actions, nor intentions (nor in general endorsements of actions) are propositional contents.

One could perhaps argue that when endorsing a course of action one endorses some content – which would be the content entailed by the relevant practical inferential entailment. This view is not very promising if one thinks that the conclusion of practical reasoning is the performance of an action (Dancy 2004; Anscombe 1957): actions do not seem to have contents (Streumer 2010). An alternative is to think of practical reasoning as ending in the formation of an intention (Broome 1999; Brandom 1994) and then assign some content to intentions. However, even if one grants that intentions have propositional contents, it is not clear that those contents can play the required role as the conclusions of suitable practical inferential entailments. Assume, for example, that the content of an intention is a claim about the conditions of satisfaction of such an intention – that is, a claim specifying the action which would fulfil it. Then, practical inferential entailments would just become theoretical inferences:

(1)

It is desirable for me to keep dry

It is raining

If it is raining, I will only keep dry by opening the umbrella

I (will) open the umbrella

The conclusion in (1) is the claim that the agent (will) open the umbrella. Therefore, (1) just says that the fact that it is raining and opening the umbrella is the only way of remaining dry gives one reasons to *believe* that the agent (will) open the umbrella. Rather than underpinning a piece of practical reasoning in which reasons are given for adopting a certain practical attitude (an intention), (1) would underpin a piece of theoretical reasoning explaining or predicting what the agent *will actually do*.

Chapter 4

Similar problems arise if one posits that the content of an intention is a claim stating that such an intention is formed by the agent – as Brandom (1994: ch. 4) proposes:

(2)

It is desirable for me to keep dry

It is raining

If it is raining, I will only keep dry by opening the umbrella

I shall open the umbrella/ I intend to open the umbrella

(2) underpins a piece of theoretical reasoning supporting the belief that the agent will form a certain intention. If the agent happens not to form the intention to open the umbrella (say, she got distracted by irrelevant considerations), the inferential entailment will be revealed to be incorrect, but the piece of practical reasoning recommending to open the umbrella will remain good: opening the umbrella is the thing to do, regardless of whether the agent does actually intend to do it. Thus, the goodness of the relevant piece of practical reasoning does not depend on the correctness of inferential entailments like (2).

Given these difficulties, one option is just to claim that practical reasoning is not underlain by inferential entailments (by practical arguments). Actions (and intentions) would be responsive to reasons, provided by facts, but these reason-giving relations would not be supported by inferential entailments between propositions (see Streumer 2010). In the same way, in section 4.2 I argued that having some (reliable) perceptual experience may entitle one to make an observation report, but I did not posit any inferential entailment by virtue of which such entitlement was gained: precisely, the point was that perceptual entitlements are gained non-inferentially. Analogously, in practical reasoning one would find that reasons (for action) are obtained in a non-inferential way (certain facts support or recommend certain actions, but not because of some inferential entailment between propositions).

The non-inferential strategy just outlined is a coherent view that deserves being explored. However, I think that there is still one plausible way of construing practical inferential entailments. The idea is to think of such inferential entailments as ending not in intentions (or their contents) but in *practical normative judgments* – that is, claims about what the agent ought to do or is entitled to do (Audi 1989; Chrisman 2008; Del Corral 2013). For instance:

(3)

It is desirable for me to stay dry

It is raining

If it is raining, I will only keep dry by opening the umbrella

It is desirable for me to open the umbrella/ I ought to open the umbrella⁶³

The conclusion of (3) has a suitable propositional, assertible form; in addition, (3) seems to underwrite pieces of practical reasoning, rather than predictions about the psychology or behavior of the agent. If (3) is a correct inferential entailment, then the premises will give reasons for believing the conclusion, that is for believing that opening the umbrella is what the agent ought to do (what is desirable for her to do). In turn, believing that the agent ought to open the umbrella (that opening the umbrella is desirable for her)

⁶³ When the relevant action is not a necessary means for achieving the goal, but just one among many permissible alternatives, the normative claims concluding the inference will specify that the agent is allowed to (albeit not obliged to) perform the action. For example:

(4)

It is desirable for me to eat something sweet

The chocolate cake in the fridge is sweet

I may eat some chocolate cake

amounts to believing that the agent has reasons to do so⁶⁴ (insofar as one has reasons to do something if one ought to do it or it is desirable for one to do it). Moreover, it is also very plausible to think that if someone correctly believes that she has reasons to do something, then she actually has reasons to do it. Thus, the inferential entailment represented by (3) seems to give reasons for the agent to endorse a certain course of action. Rather than entailing that the agent will actually perform some action (or form some intention), (3) entails that some action ought to (or may) be performed.

One may protest that (3) is not a practical inferential entailment, but rather a theoretical one, since it offers reasons to *believe* some (normative) claim. This may be so, but, still, the resulting beliefs have a direct role in motivating and justifying action – in deciding what to do. In this sense, it is not misleading to say that the inferential entailment is practical and has a practical conclusion.

In a Brandomian spirit, one can see the normative vocabulary involved in the conclusions of (3) as performing a (logical)-expressive function (in line with the discussion in section 4.3.3): this vocabulary would allow us to speak explicitly about the practical commitments and entitlements of an agent – by means of assertions which, in turn, can be explicitly challenged and justified. Before the introduction of such vocabulary, treating an agent as having certain practical commitments and entitlements was something that could only be done implicitly in practice. Once this expressive normative vocabulary is available, it becomes possible to specify explicitly the inferential entailments underlying pieces of practical reasoning, and to subject such entailments to critical scrutiny.

To recap: when engaging in practical reasoning one takes certain facts as recommending or supporting a certain course of action – that is, as offering reasons for endorsing such a course of action. It is not clear whether these reason-giving facts support the endorsement of the action by virtue of standing in inferential entailments with other facts (as would happen in the case of theoretical reasoning). Inferential entailments between propositions seem to give reasons to *believe* or accept the entailed proposition, on the basis of the facts entailing it. Bearing this in mind, I have argued that

⁶⁴ Or to do something else that is a good means for achieving such an end (e.g. taking the umbrella from its case), as discussed in section 4.3.3.

the most promising way of characterizing practical inferential entailments is to think of such entailments as concluding in a normative claim – a claim about the practical obligations and permissions of the agent. These entailments would provide reasons to *believe* that one has certain practical obligations or permissions. However, these beliefs would have immediate consequences for deciding what to do, so that the inferential entailment can be used to offer direct support to pieces of practical reasoning (reasoning about what to do).

In the next section I will discuss the relation between the success of intentional actions and the goodness of the pieces of practical reasoning that licensed such actions.

4.4 Intentional agency and success

Intentional actions can be thought of as actions with a goal. In the previous sections, I have discussed one way in which such actions can be normatively assessed: one can ask whether an agent is entitled to perform a given action – whether there are good reasons for her to perform it. Intentional actions can also be evaluated with respect to their success – more specifically, their success in fulfilling the goal of the agent. In this section, I will argue that these two dimensions of assessment are strongly related: if an action is not successful, it was not properly entitled (the agent cannot have proper entitlement to perform it as a sufficient means for achieving her goal).

One first way in which an agent's performance may fail is by not succeeding to perform the action intended. Imagine, for instance, that I have reasons to open the window (the room is very hot), so I form the intention to open it. If I actually manage to open the window, I will have succeeded in doing what I intended. In contrast, if the window panel is stuck and I cannot move it, I will have failed. In Brandom's terminology, I will have failed to discharge the practical commitment I undertook when forming my intention.

This is not, however, the only way in which an action can fail. The performance of an intended action can also fail in fulfilling the goal of such an action – that is, the goal in light of which the action was recommended. For example, if by opening the window the

room actually gets cooler, I will have achieved my goal of making the room less hot: my action will be successful. By contrast, if the room remains very hot despite the window having been opened (maybe the air outside was even hotter), my action of opening the window will have failed in achieving its goal.

I will focus on pieces of practical reasoning in which some action is recommended as a sufficient means for achieving some goal (rather than as an action that merely increases the likelihood of achieving its goal).⁶⁵ When the performance of an action is recommended by a good (valid and sound) piece of reasoning of this sort, it cannot fail in achieving its goal.⁶⁶ If the agent manages to perform the action intended and, nonetheless, she does not achieve her goal, then the piece of practical reasoning recommending the performance of the action was not actually good: there were not actually good reasons supporting the performance the action. After all, if the action fails in achieving the goal, it will turn out not to be a sufficient means for securing such a goal – contrary to what the piece of reasoning made it appear.

Typically, means-end pieces of practical reasoning (of the sort I am considering) will include among their premises a conditional premise stating that the action is a sufficient means for achieving the agent's goal – provided that the rest of the premises hold. Think, for example, of the following piece of reasoning (I will present the inferential entailment associated with the relevant pattern of reasoning):

(5)

It is desirable for me to drink the tea without getting burnt

⁶⁵ In the same way, I have been focusing on theoretical reasoning that licenses full belief in a proposition – rather than probabilistic reasoning which only assigns a certain degree of likelihood to some claim.

⁶⁶ Of course, the validity of inference does not need to be deductive. I will say that an inference is valid if and only if the considerations presented as premises constitute (whenever they are facts) good reasons for endorsing the conclusions – in other words, if and only if entitlement to the premises provides entitlement to the conclusions. A valid but unsound inference would be one in which the premises are not facts, but merely considerations that would provide good reasons for the conclusions if they were facts.

The tea is merely warm

If the tea is merely warm and I drink it, I will not get burnt

I may drink the tea

The means-end conditional premise, plus the second premise “The tea is merely warm” (which states that the antecedent of the conditional holds), ensures that, whenever the action fails to achieve her goal, some of the premises will not actually obtain – so that the piece of reasoning is unsound. If the non-conditional premise “The tea is merely warm” obtains, but nevertheless the agent burns herself when drinking the tea, then the antecedent of the conditional premise will hold without the consequent doing so; therefore, the conditional premise will be false and the piece of reasoning unsound.

Perhaps the presence of such means-end conditional premises is not always required in order for a piece of practical reasoning to be non-enthymematic (in the same way that inferentialists allow for good material inferences without a conditional premise stating that the conclusion of the inference follows from its premises). For instance, the premises in (6) below could be seen as offering reasons recommending the action of opening the umbrella, without needing an additional conditional premise stating that opening the umbrella is a sufficient means for staying dry under the rain.

(6)

It is desirable for me to stay dry

It is raining

I should open the umbrella

It may be that the action of opening the umbrella fails to achieve the goal of remaining dry even if all the premises of (6) hold: perhaps the umbrella had holes, or the wind made the rain get under the umbrella. What happens in these cases, I submit, is that the inferential entailment represented by (6) stops being correct (being valid). (6) is a defeasible, non-monotonic valid inferential entailment: there are undercutting defeating conditions (such as the umbrella having holes) that make it incorrect.⁶⁷ When such undercutting defeating conditions obtain, the premises of the inference do not actually provide good reasons for endorsing the conclusions: that it is raining is not a reason for opening the umbrella, if the umbrella is full of holes.

Thus, the introduction of a means-end conditional premise, like “If I open the umbrella, I will remain dry”, turns defeasible, non-monotonic practical inferences into indefeasible, monotonic ones: those conditions that were undercutting defeaters for the non-monotonic inference will now make the monotonic version of the inference unsound. In any case, failures in achieving the agent’s goal by performing the recommended action will reveal that the action was not actually supported by a good piece of practical reasoning: either it was invalid (perhaps it was defeated), or it was unsound.

In the next sections I will discuss how the assessment of the success of the actions of an agent may lead to the update and revision of her views. I will start by presenting Brandom’s (2008) account of this process of updating. After that I will introduce my own proposal, which will be developed in chapters 7 and 8.

4.5 Brandom’s TOTE-cycles

In *Making it Explicit*, Brandom barely discusses the place occupied by evaluations of practical success in the games of giving and asking for reasons. It is not until his more recent work, in particular *Between Saying and Doing*, that the notion of practical success

⁶⁷ I will discuss these issues in more detail in chapter 7.

(i.e. success in achieving the goals of intentional actions) plays a more important role – but even there it remains implicit for the most part.

The notion of success is introduced in *Between Saying and Doing* – or at least its introduction is suggested – in the context of the discussion of feedback cycles of perception and action. Brandom (2008: lecture 6; also 1994: 331-33) calls these cycles Test-Operate-Test-Exit cycles (TOTE-cycles); they are open sequences of action and perception, in which the evaluation of the results of a performance is used as input for further performances. So, in a TOTE-cycle the agent would perform an action, then she would observe and evaluate its outcomes and subsequently this evaluation would lead to further actions.

Brandom takes the notion of TOTE-cycles from the theory of control systems, in particular from the work of Miller, Galanter and Pribram (1960) – although feedback driven goal-oriented behavior had already been studied in the field of cybernetics by Rosenblueth, Wiener and Bigelow (1943). Moreover, perception-action feedback cycles play a central role in Gibson's ecological psychology (1979), and also in sensory-motor (O'Regan and Noë 2001; Noë 2004) and enactivist (Hurley 2001; Gangopadhyay and Kiverstein 2009) approaches in the cognitive sciences. The original idea of perception-action feedback loops was put forward by Dewey (1896), and is one of the key notions of classical pragmatism.⁶⁸

So, there is a long tradition that explains goal-directedness in terms of feedback controlled behavior – in particular, in terms of perception-action loops. Brandom's appeal to TOTE-cycles can be included in this tradition. However, Brandom only presents his proposal in a sketchy way, and it is not clear how it should be integrated with the rest of his views. To be in a position to discuss how the introduction of TOTE-cycles bears on Brandom's project of accounting for objective representation, it is necessary to have a much more detailed description of the role these cycles play in discursive practices, which Brandom does not provide.

⁶⁸ Dewey writes:

What we have is a circuit, not an arc or broken segment of a circle. This circuit is more truly termed organic than reflex, because the motor response determines the stimulus, just as truly as sensory stimulus determines movement. Indeed, the movement is only for the sake of determining the stimulus, of fixing what kind of a stimulus it is, of interpreting it. (Dewey 1986: 363).

In particular, Brandom does not specify explicitly what the evaluation of actions consists in, and he does not develop in detail how such evaluations are related to the updating of the agent's views (i.e. her discursive commitments). Actually, the example that Brandom analyzes more carefully (2008: lecture 6), is an instance of expansion of observation, rather than a genuine TOTE-cycle. In a process of expansion of observation, one first observes something and extracts some inferential consequences of the observation; then one performs another observation and checks if this second observation is compatible with the consequences derived from the first one. The role played by action is much less relevant – at least overtly relevant – than in TOTE-cycles.

The example presented by Brandom is the following: some substance tastes sour, so (in accordance with the inferential norms of the community considered) it is classified as an acid; but it is also observed that the substance turns phenolphthalein red, which for that community is incompatible with the previous classification. Even if this process could be rephrased as a TOTE-cycle, it is not described in those terms as it stands. It would seem that we just have two observations – that may give rise to incompatibilities in our beliefs –, rather than perception followed by action, evaluation of the results of the action and further performance. Actually, the agent does not need to perform any action in order for the second observation (that the substance turns phenolphthalein red) to take place: perhaps the agent just happens to observe, passively, how phenolphthalein reacts to the substance. In particular, there is no mention in Brandom's picture of any action being evaluated as successful or failed – and of such evaluations leading to updates in the agent's beliefs.

Brandom's focus (at least since *Between Saying and Doing*) on practical interactions with the world goes in the right direction, I think, towards offering an account of representational practices. In particular, the idea of perception-action feedback loops, which is at the heart of pragmatist views of representation, should play a crucial role in such an account. However, his proposal is too schematic and needs to be fleshed out; moreover, Brandom would need to show how he can overcome several difficulties that plague projects similar to his (among them, objections to those projects which have been voiced by Brandom himself, as I will discuss in chapter 5).

In the next section I will present my own framework for examining how the agent's views are updated in light of the assessment of the success of her actions. In chapter 7 and 8 I will develop this framework so as to offer an account of the role of

perception and action in discursive practices that vindicates the sensitivity of such practices to the way the world is – in particular, an account that manages to address the sort of difficulties I will discuss in chapters 5 and 6.

4.6 Practical success and the updating of representational views

I have argued that the performance of actions for which the agent has proper entitlement cannot fail in achieving the goals of such actions. In this way, whenever an action fails in fulfilling its goals, the agent was not properly entitled to performing the action as a means for her goal. Thus, the agent did not actually have good reasons for doing what she did: there was something bad in the piece of reasoning that she took to offer entitlement for the performance of her action. This means that the agent is endorsing commitments she is not entitled to: maybe she is endorsing as correct an inferential pattern that is actually incorrect, or she is committed to accepting certain premises that do not actually obtain (and which it is therefore incorrect to accept). The evaluation of the action as failed will be incompatible with the previous commitments of the agent, so she will need to update her views in order to solve this incompatibility.

Let me present an example. Imagine that I have the goal of drinking my tea without getting burnt. Now, I also believe that the tea is not too hot, and I regard this fact as giving me reasons to drink the tea, without fear of being burnt. My confidence in this piece of practical reasoning is underpinned by my endorsement of the inferential entailment (5):

(5)

It is desirable for me to drink the tea without getting burnt

The tea is merely warm

If the tea is merely warm and I drink it, I will not get burnt

I may drink the tea

Now, suppose that I drink the tea and, while doing so, I feel that I am getting burnt. This is to say, I have a perceptual response that entitles me to form the belief that I am getting burnt. If this is the case, I will be entitled to assess my action as having failed in achieving its goal. Such a failure is incompatible with the truth of the consequent of the means-end conditional embedded in the premises of (5). However, I am committed to the truth of the antecedent of such a conditional: I thought that the tea was not hot and I also take myself to have actually performed the recommended action (drinking the tea). So, I am also committed to accepting the truth of the consequent of the means-end conditional premise (i.e. that I would not burn while drinking). This is incompatible with my perceptual experience of being burnt (which entitles me to thinking that I am actually getting burnt).

To remove this incompatibility, I have to choose between several options. First, some of the antecedents of the means-end conditional premise could be actually false (and my commitment to them, thereby, incorrect).⁶⁹ Second, the conditional premise may be itself false: it may encode a bad inferential entailment. This would happen if the theoretical inference encoded by such a conditional had been made bad by some ignored undercutting defeater. For instance, if the tea mug was very hot, then the fact that tea is merely warm does not entail that I will not get burnt when drinking (the mug may burn me). Third, my evaluation of the action as failed may be incorrect: perhaps my perceptual

⁶⁹ The premise stating that the agent's goal is something desirable could also be false, but this would not affect the incompatibility resulting from my burning while drinking the tea: it would only make the goal of my action a goal not worth pursuing.

It could also be the case is that the agent's goal is impossible to attain (e.g. the goal to turn gold into water, or to make some particle travel faster than light), Note, however, that even in these cases the failure of the agent's action in securing her (impossible) goal would reveal that there is a mistake either in the agent's beliefs or in the inferences she endorses. If the agent took herself to be entitled to performing the action it was because she thought that (given her other beliefs) it was a good means for achieving her (impossible) goal. But, as a matter of fact, no possible action is actually a good means for achieving the agent's goal – remember, it is an impossible goal. Thus, the agent would be mistaken in thinking that such an action is a good means for achieving her goal.

experience of getting burnt was not accurate and did not actually entitle me to think that I was getting burnt.

Once the source error is assigned to one of these factors, I may update my views accordingly and engage in a new process of practical reasoning. For instance, if I conclude that my belief that tea is not too hot is false, I ought to withdraw my commitment to accepting such a claim. This may lead to the performance of new actions, such as blowing and stirring the tea so as to make it cool down faster. These sequences of action and observational evaluation give rise to open-ended perception-action feedback loops, in which the agent revises and updates her representational views in accordance with the assessment of the outcomes of her interactions with the world.

In conclusion, when an agent assesses as failed an action that she took to be properly entitled, she will find herself endorsing incompatible commitments. Therefore, she will be required to revise her views. On the face of it, it might seem that the failure of the action must be due to the falsity of some of the non-conditional premises (“The tea is not too hot”, in this example). However, we have just seen that the situation is more complex: there are other possible sources of error. As a consequence, the pragmatic test of the success or failure of our actions is not as precise as we might think: we cannot isolate the truth or falsity of our non-conditional premises just by looking at the success of the actions they prompt. But, on the other hand, this imprecision makes the pragmatic test more resilient. If we did not take into account all the extra possible sources of practical failure, we would leave unexplained those cases in which the action did not succeed although the non-conditional premises are all true. In particular, ignored facts (such as the mug being very hot) can thwart the success of our actions, even if our endorsement of the non-conditional premises is properly entitled. I will discuss these issues in detail in chapter 7.

4.7 Conclusions

In this chapter, I have discussed the way in which discursive practices are connected with perception and action (taking Brandom’s model as my starting point, although I have departed from it in several places). On the one hand, reliable perceptual experiences may entitle agents to endorsing certain claims – to making certain observation reports. On the

other hand, entitlement to some claims may provide entitlement to performing certain actions: the facts expressed by those claims would give reasons for performing those actions, as a means to fulfilling the agent's goals.

I have worked with stringent notions of perceptual entitlement and of entitlement to acting. According to these stringent notions, one cannot be perceptually entitled to endorsing false beliefs; likewise, a properly entitled action cannot fail in fulfilling its goal. So, whenever an agent evaluates that her action failed, either her action was not properly entitled or her evaluation is incorrect. The detection of practical failures, thus, leads to the revision of the agent's commitments.

The resulting picture is the following: on the one hand, by virtue of their representational content, the agent's beliefs license the performance of certain actions (in light of the agent's goals). But, on the other hand, the outcomes of these actions (their success or failure) reveal whether the agent's views were correct – whether they provided proper entitlement for the performance of those actions. Thus, the correctness of discursive moves (i.e. of endorsing certain views) is dependent on the outcomes of the agent's interactions with the world. If these outcomes are sensitive to the way the world is, then it can be concluded that the correctness of discursive moves answers to the way things are in the world. In this way, human representational practices would be shown to incorporate an objective dimension.

In chapters 7 and 8 I will develop my account of representational practices in more detail. Before doing so, in chapters 5 and 6 I will discuss projects with similar aims to my own – in particular, success semantics and etiological accounts of biological goals. I will identify the most pressing difficulties faced by these projects, in order to discuss in chapter 7 and 8 how my proposal can deal with such difficulties. More specifically, I will consider two main problems: the possibility of ignored obstacles and the need to determine objective conditions of success

Chapter 5: Success Semantics

5.1 Introduction

Success semantics is a name usually associated with a group of theories about the representational content of beliefs and sentences. In a straightforward version of success semantics (Whyte 1990; Dokic and Engel 2002), the truth conditions of a belief would be those that guarantee the success of every action derived from that belief in combination with any possible desire. There are more sophisticated formulations (Blackburn 2010: 181-199; Nanay 2013), but they all find a common theme in the exploration of the relations between the truth conditions of beliefs and the practical success of the actions prompted by such beliefs.

Despite its name, success semantics is really a meta-semantical theory; it is a theory about what it takes for a belief to have certain semantic content, characterized according to some semantic theory, typically truth-conditional. However, for convenience I will retain the label 'success semantics', which is common in the literature.

On the face of it, success semantics seems a promising approach for developing a pragmatist meta-semantics (i.e. the sort of meta-semantics characterized in 2.3, according to which expressing semantic contents is a matter of playing a certain role in a practice). According to success semantics, a belief counts as having certain truth conditions if it leads to actions with certain success conditions. In this way, what it is to express a semantic content with some truth conditions is explained in terms of use, of practical role: it is to be potentially involved in the production of actions that are successful in certain conditions but not others. Attributions of truth conditions have, according to success semantics, a clear practical significance. In this chapter I will discuss whether success semantics actually offers a viable practice-based account of what it is to have semantic content (that is, a viable practice-based meta-semantics). I will present different formulations of success semantics and discuss some of the objections they have received. My conclusion will be that, in its available formulations, success semantics is not in a position to offer satisfactory replies to such objections. However, it is interesting to examine the shortcomings of this sort of view, since they will reveal challenges that my own proposal has to face. In following chapters, I will propose a

pragmatist-inferentialist account of representation which manages to overcome – or at least tame – the problems undermining success-semantics.

The structure of the chapter is as follows. In section 5.2 I introduce success semantics. In section 5.3 I discuss whether truth is necessary for success. After that, in section 5.4 I discuss whether truth is sufficient for success; in particular, I examine the problems posed by false collateral beliefs and by the possibility of ignored obstacles. In section 5.5 I reassess the question of whether success is necessary for success, in light of the issues discussed in 5.4. In section 5.6 I critically examine Blackburn's version of success semantics, which tries to avoid the problems discussed in the previous sections. Finally, in 5.7 I introduce a further problem for success semantics, the determination of success conditions.

5.2 Success semantics

One of the formulations of Peirce's pragmatist maxim is:

The entire intellectual purport of any symbol consists in the total of all general modes of rational conduct which, conditionally upon all the possible different circumstances and desires, would ensue upon the acceptance of the symbol.

(Peirce 1998: 346)

An obvious way of developing this pragmatist insight is to suggest that the truth conditions of a belief (or proposition) may be identified by looking at the conditions of success of the actions that are derived from it. When we accept a proposition, certain lines of action become recommended (in view of the goals of the agent); the success or failure of those recommended actions will tell us if it was appropriate to accept the proposition – if the proposition was true.

Ramsey sketched such an account in 'Facts and Propositions'. Ramsey endorsed a pragmatist analysis of meaning and content: 'the meaning of a sentence is to be defined

by reference to the actions to which asserting it would lead, or, more vaguely still, by its possible causes and effects' (1927: 170). If one now wants to identify the representational purport of such pragmatist content – the truth conditions associated with it –, it is quite natural to try to work out some form of success semantics. This is what Ramsey did:

Thus any set of actions for whose utility p is a necessary and sufficient condition might be called a belief that p , and so would be true if p , *i.e.* if they are useful (Ramsey 1927: 159).

Ramsey advanced this proposal in connection with the kind of beliefs that we can attribute to non-linguistic animals – that is, beliefs that are expressed through the behavior of agents, rather than by their use of words or symbols. However, Ramsey's suggestion clearly gestures towards a more general account of representational content. The theories that have tried to develop this sort of account are usually gathered under the label of 'success semantics.'

Note, as explained above, that in spite of this label success semantics is a theory at the level of what I called in 2.3 meta-semantics. In the previous quote, Ramsey puts forward a view explaining what makes something a belief with semantic content p – while a semantic theory would characterize the content p , by assigning it semantic values, for instance truth-conditions.

Whyte (1990) proposed one of the first worked out versions of success semantics.⁷⁰ Dokic and Engel (2002) provide a very similar version of the thesis put forward by Whyte, which they call 'Ramsey's Principle':⁷¹

⁷⁰ This is how Whyte synthesizes his proposal:

A belief's truth condition is that which guarantees the fulfillment of any desire by the action which that belief and desire would combine to cause. (Whyte 1990: 150).

Ramsey's Principle (RP): A belief's truth conditions are those that guarantee the success of an action based on that belief whatever the underlying motivating desires. (Dokic and Engel 2002: 46).

The core idea of success semantics, thus, is that the truth conditions of beliefs can be derived from the conditions of success of the actions that follow from such beliefs – in combination with suitable desires. So, for instance, the belief that there is water in the flask will lead me to drink from it, in order to quench my thirst. If my belief is actually true, and there is water in the flask, I will succeed in quenching my thirst; however, if my belief was false and the flask was empty, it is most likely that I will fail.

In this chapter, I will explore different ways of developing success semantics, and discuss some of the objections that it has to face.

5.3 Is truth necessary for success?

If RP is compared with Ramsey's original proposal, one sees that Ramsey argued that the truth of a belief was a *necessary and sufficient* condition for the 'utility' of the actions associated with it, while the reformulated Ramsey's principle (RP) only talks of *sufficient* conditions: the truth of the belief *guarantees* the success of the actions that follow from it. This amendment allows success semantics to escape a first obvious objection: an action may be successful even if it is based on false beliefs. Perhaps several false beliefs

Whyte, moreover, favours a causal functionalist characterization of beliefs: a belief would be identified by its causal properties (what actions it causes when combined with different desires). This is not the only possible functionalist position: for instance, beliefs could be characterised by the actions that *ought to* – or *may* – follow from them (in combination with suitable motivational states). In what follows, I will not be concerned about this distinction – except where it is specifically relevant.

⁷¹ See also Mellor (1991) and Papineau (1987).

combine felicitously so as to bring successful actions; or perhaps there are other facts in the world that explain our practical success, despite the falsity of our beliefs. For instance, if I believe that the train to Oxford is at 4 o'clock, when in fact it sets off at 3, but at the same time my watch is one hour fast, I will successfully take the train to Oxford. Similarly, if I mistakenly think that the liquid in my flask is orange juice, when it is actually water juice, I will still be able anyway to quench my thirst by drinking it (because water also quenches thirst).

However, does not success semantics lose its interest with this modification? If truth guarantees practical success, but is not required for it, then the conditions of success of an action will not identify the conditions of truth of the beliefs that lead to it – because the action may still be successful when such conditions of truth are not met. Both Whyte (1997) and Dokic and Engel (2002) reply that the truth of a belief might not be necessary for the success of a particular action, but, nevertheless, it will be typically necessary for the success of the *totality* of the actions it may lead to. Remember that RP demanded the success of every action based on the belief under consideration, in combination with 'whatever' desire.⁷² Some of these actions will fail if such belief is false.⁷³ For example, my wrong belief that I have orange juice in my flask (when I actually have water) may allow me to quench my thirst, but other possible actions – such as treating somebody who loves orange juice but does not care about plain water – would fail.

⁷² Blackburn (2010: 186-187) agrees that the strategy of considering all possible actions derived from a belief is promising, but he is afraid that such strategy distances success semantics from the actual empirical evidence – since many of these actions will never be performed. This problem is analogous to the old asymmetry between falsifiability and verifiability in generalizations or scientific theories: if it does not worry us that we can never finish to verify a scientific theory (while a single anomaly falsifies it), we should not worry here, either – even if we grant that there is a potentially infinite number of possible actions derivable from a given belief.

⁷³ Is it possible that the world conspires in such a sophisticated way that all the actions possibly entailed by my false belief are successful (all the actions that may be performed by my and also by others, on the basis of that belief)? Well, according to success semantics this situation would be indistinguishable from situations in which my belief is true: there would be no reason to say that it was a false belief.

I will grant that truth is necessary for the success of all possible actions following from a given belief. In the following section I will discuss the view that truth is sufficient for success. I will argue that this view faces serious problems.

5.4 Is truth sufficient for success?

5.4.1 False collateral beliefs

The truth of a belief, as we have seen, is not always necessary for the success of a particular action that follows from it – although it seems to be necessary for the success of *all* such actions. A further question is whether truth is at least *sufficient* for success. Without some qualifications, the answer seems to be also negative. Clearly, the truth of a single belief is not enough to guarantee the practical success of all the actions that may derive from it. The reason for this is that a belief usually leads to an action in combination with other beliefs. So, my belief that the train to Oxford sets off at 4 o'clock, will only make me go to the station if I also believe that now is almost 4 o'clock. And, obviously, my action may fail because some of these collateral beliefs are false (if my watch is late, I will probably miss the train, even when I was a right about its departure time).

Thus, a true belief, combined with a false one, may lead to failed actions (Brandom 1994b). One possible reply is to try to focus on actions produced by a single belief (Mellor 1991; Nanay 2013). However, it is doubtful that such a limited range of actions is enough to characterize the complete conditions of truth of beliefs. Arguably, some of these conditions will only be revealed through the conditions of success of actions derived from complexes of beliefs – at least if we do not endorse an atomistic view of mental content. After all, most actions only follow from sets of beliefs.⁷⁴

Rather than restricting the analysis to actions produced atomistically, an alternative solution for this problem is to say that RP gives us conditions of truth for the whole set of

⁷⁴ One could argue that, actually, no action follows from a single belief. To be fair, Nanay (2013) does not focus on full-fledged beliefs, but on a more basic kind of representation, which he calls 'pragmatic representations'. So, he is not claiming that beliefs are connected atomistically with action, but only that pragmatic representations are.

beliefs that combine to produce the action (Whyte 1990; Dokic and Engel 2002; Brandom 1994b). We would be able to isolate the truth conditions of individual beliefs by contrasting the truth conditions of different sets of beliefs. According to a first strategy, the truth conditions of the individual belief would be found by comparing the truth conditions of various sets which only share that individual belief (Whyte 1990; Dokic and Engel 2002); in a second strategy, the individual truth conditions would be identified by contrasting the truth conditions of sets that only differ in that individual belief (Brandom 1994b).

For example, my belief that it is almost 4 o'clock will make me go to the station, when combined with the belief that the train to Oxford leaves at 4 and the desire to catch that train; on the other hand, it will make me go to the cinema if I also have the belief that a film I want to watch starts at 4. According to the first strategy, I would isolate the conditions of truth of the belief that it is almost 4 o'clock by contrasting the conditions of success of sets of beliefs like the ones just presented. The second strategy would look instead at the contribution made by a given belief to sets in which it did not figure previously

One first point should be made. Whyte thinks that, in order to isolate the individual truth conditions of some belief, it is enough to compare the truth conditions of *two* conjuncts sharing only that belief.⁷⁵ However, this is not generally right. In general, from the disjunction of the truth conditions of "*p & q*" and the truth conditions of "*p & r*", it is not possible to isolate the truth conditions of "*p*" (unless "*q or r*" is a tautology). The same difficulty arises with the method of contrasting the truth conditions of *two* sets differing only in a single belief: the truth conditions of "*p*" cannot be derived, in general, from the truth conditions of "*p & q & r*" and "*q & r*". It is possible to extract individual truth conditions out of conjunctions, but not by magic: in general, we will have to resort to *more than two* conjunctions.⁷⁶ So, focusing in Whyte's method of comparing

⁷⁵ 'How can we pick out the part of the A2's success condition to which B2 answers for its truth (namely, there being chocolate in the fridge)? Very simply. Take an action caused by B2, but not B3 and B4, and see what that action's success condition shares with A2's success condition.' (Whyte 1990: 151).

⁷⁶ So, imagine that we know that the conjunct "B1 and the house is small" is true whenever the house is red and small, and that the conjunct "B1 and the house is empty" is true whenever the house is red and empty.

conjunctions sharing only the interesting belief, we will get closer to those individual conditions if we keep uniting (through disjunctions) conjunctions involving that belief. As Dokic and Engel say: “The truth conditions of the belief is the invariant of all the truth conditions of *all the sets* to which it belongs” (2002: 49; my emphasis).

In general, the truth conditions of a belief “B1” may be singled out from the truth conditions of several conjuncts sharing “B1” when the disjunction of all the other components of the conjuncts is true in all circumstances. So the disjunction for “(B1 & B2)”, “(B1 & B3)” ... “(B1 & Bn)” is: “[B1 & (B2 or B3...or Bn)]”. If “(B2 or B3...or Bn)” is true in all circumstances, then the truth conditions of B1 are given by the truth conditions of the whole disjunction. This result points to a general strategy for detaching individual truth conditions. If one has the truth conditions of a given conjunction “(B1 and B2)” – say, “There is only one open door in the house” and “The door in the kitchen is open” –, the next step would be to look for the truth conditions of a conjunction “(B1 & complementary-B2)” – where by “complementary-B2” I mean a belief such that its disjunction with B2 is true in all circumstances). So, the second belief would be “The door in the kitchen is closed”. In that case, “(B2 or complementary-B2)” will be true in all circumstances – the kitchen door is either open or closed –, and the truth conditions of “B1” may be extracted from the disjunction of both conjunctions.

What are the truth conditions of the claim “B1”? One would be tempted to answer that “B1” is the claim “The house is red”. But – among many other possibilities – it could also be the claim “The house is red and (small or empty)”. Likewise, if the conjunct “The house is B1, small and empty” is true when the house is red, small and empty, “B1” may be the claim “The house is red”, but also (for instance) the claim “The house is red or (yellow and big)”.

It could be argued that this sort of indeterminacy only arises because I am letting beliefs refer to any sort of properties, including wildly disjunctive ones (like the property of being red or (yellow and big)). Perhaps the indeterminacy could be solved if one follows Lewis (1983) in privileging some class of properties as more ‘natural’ (e.g., as associated with ‘universals’). Then, it could be said that beliefs should be interpreted as referring to the most natural properties available (being red would be more natural than the disjunctive property: being red or (yellow and big)). In particular, properties relevant for action planning and production would tend to be natural. I will not pursue this line here, however, in part because I am doubtful about the project of privileging certain properties as natural or basic.

A further possible problem for these comparative methods – one might think – is that there are some conditions for the success of actions which are shared by all – or almost all – the actions that can be carried out in normal contexts. For instance, the success of our actions usually depends on the existence of oxygen in the environment, or on the effects of gravitational forces. These general 'stability conditions' (Whyte, 1990), however, would not be part of the individual truth conditions of the specific beliefs that contribute to produce the action. If we cannot find actions whose success is independent of such 'stability conditions', it would seem that we are not going to be able isolate the truth conditions of individual beliefs – they will always be merged with these general conditions for the success of actions.

However, at first sight there seems to be no reason why these background stability conditions cannot be factored out applying either of the detachment methods introduced above. Most 'stability' conditions will not affect every possible action, but only every action actually performed by humans, or actually performed in normal contexts (say, on the Earth, at usual temperatures etc). Therefore, it seems that, in order to get rid of the background stability conditions, the only thing one has to do is to look for alternative actions that are not under the influence of such conditions. Whyte (1990) and Dokic and Engel (2002) argue that this is in general possible, because we can appeal to the success of actions under counterfactual circumstances. So, even if all our actions on the Earth are performed in the presence of gravitational forces, we can always consider what would happen if things were different – if there were no gravitational forces operating. And actually, some of these actions beyond standard background stability conditions can be sometimes carried out physically, not just as mental experiments: for example, scientists in a spaceship can actually test the outcomes of actions performed in the absence of gravity.

However, even if we grant that the methods of detachment introduced above are successful, there is still the question of whether agents have beliefs – either explicit or implicit – about all these background conditions. It seems implausible to attribute such background beliefs (about gravity, oxygen in the atmosphere etc) to agents, even when they perform simple tasks; however, if we do not do so, RP will be violated: in situations in which one of these background stability conditions does not obtain, we would have a failed action which was not based on any false belief.

In the next section, I will examine how the sort of issues raised in the last paragraph may threaten the viability of success semantics. More specifically I will discuss the problem posed by the possibility of ignored obstacles to our practical plans.

5.4.2 Ignored obstacles

I have been arguing that the truth of a belief may not guarantee the success of the actions that follow from it, when collateral beliefs are required in order to derive such actions. The success of an action could only be guaranteed, therefore, by the truth of all the beliefs that lead to it. Nevertheless, even if all our beliefs are true, ignored facts may still thwart our practical projects (Brandom 1994b). Imagine I correctly believe that I poured water in my flask; I may still fail in quenching my thirst by drinking from the flask, if I am unaware that there is a hole in it. The absence of each of these possible obstacles would constitute further conditions for the success of the action. One of the conditions for my success in drinking from the flask is that the neck is not frozen, or that the cap is not blocked. These background (or 'anti-obstacle') conditions for the success of the action of drinking from my flask are not, however, part of the truth conditions of the belief that there is water in my flask.

The advocate of RP would have to show that ignorance implies some sort of false belief or mistake. It can be argued that, when I intend to drink from the flask so as to quench my thirst, I have the belief (perhaps implicitly) that there are no holes in the flask. If this is so, when as a matter of fact there are holes, one of my beliefs would turn out to be false, so RP would be able to account for the failure of my action. The problem is that, at least according to Brandom (1994b), there are an indefinitely large number of possible ignored obstacles for the success of my action. Perhaps the cap of the flask is stuck; or there is some substance inside that makes the water undrinkable. Arguably, with a bit of imagination one will always be able to find new ways in which actions may fail due to ignored facts (Brandom 1994b). Therefore, it is doubtful that we will be able to provide a list of all these facts. And, even if we are able, it is not realistic to demand from agents to have beliefs about all those facts: such a demand would be too psychologically burdening (Dokic and Engel 2002, 2005) – at least, as it will be seen below, if these beliefs are taken to be explicit or actually formed by agents. The problem is not only that the number of

required beliefs is indefinitely large, but also that many of those beliefs will be rather bizarre and convoluted, and are not likely to have ever crossed the agent's mind (think of beliefs about the possibility of evil wizards casting defeating spells).

I will discuss two ways of defending RP against the problem of ignored obstacles – proposed by Dokic and Engel (2002, 2005) and by Whyte (1990), respectively. After examining their shortcomings, I will present an alternative version of success semantics, by Blackburn (2010: 181-199), in which truth is not taken to guarantee success.

5.4.2.1 *Dokic and Engel's proposal: beliefs implicit in action*

Dokic and Engel (2002, 2005) have put forward a way of trying to deal with the problem of ignored obstacles. As I have just pointed out, they think that it is not plausible that agents have explicit beliefs about every possible impediment for the success of the action; however, they argue that such beliefs may be *implicit* in an agent who successfully performs the action.

Dokic and Engel work with a peculiar notion of implicit belief, which they derive from debates in epistemology. According to this view, our successful performance of an action puts us in a position to acquire the knowledge that any impediment for this action has not taken place. So, my action of drinking from the flask puts me in a position to know that there were no holes in it. This does not mean that I have to form such belief in order to be able to perform the action; rather, the idea is that, were I to form such belief, I would be justified in doing so by my experience of successfully performing the action. In this sense, Dokic and Engel argue that, when drinking from the flask, I have implicit knowledge (thereby, an implicit belief) that there are no holes in it (and also that the cap is not stuck, that the mouth of the flask is not blocked etc).

One could complain that this sort of implicit belief does not play the guiding role that, from a pragmatist point of view, would be expectable from beliefs genuinely attributable to agents – many times, those implicit beliefs will not have any effect whatsoever in the behavior and mental life of the individual. Even if we grant that such implicit beliefs may be attributed to agents, I will try to argue that this does not save success semantics from the problem of ignored facts. The reason for this is that an agent

would only have these implicit beliefs when her action is *successful* – remember that it was such success that entitled the agent to infer the different particular no-impediment beliefs. However, what would be needed for a defence of RP is that those no-impediment beliefs are possessed by the agent (maybe implicitly) when her action *fails*: in that case, the no-impediment beliefs would be false, and RP would correctly predict the failure of the action – since the agent, after all, had some false belief. But Dokic and Engel cannot argue that agents have the various no-impediment beliefs *when the action fails*, not even in an implicit way. The no-impediment beliefs cannot be inferred from the failure of the action, so they would not be implicit – in the sense of implicit favoured by Dokic and Engel – when the agent’s actions fail. On the contrary, the failure of the action would at best entitle the belief that *there was some impediment* (which would be a true belief).

Dokic and Engel’s argument, therefore, does not show that agents have some false belief whenever their actions fail due to the presence of some ignored obstacle. From my failure in drinking I cannot infer that there were no holes in the flask, and therefore, I cannot appeal to the falsity of such implicit belief in order to explain the failure of my action. Perhaps Dokic and Engel may reply that they only wanted to show that, when the agent has all the relevant beliefs – or at least is in a position to form them with entitlement –, the success of her action is guaranteed.⁷⁷ But, what would happen then with the cases of practical failure due to ignored facts? In these cases, it would seem that there is no good reason for attributing any false belief to the agent, and nonetheless her action fails. Dokic and Engel, therefore, are just not addressing the problem of ignorance – and, therefore, they cannot defend the view that truth guarantees success.⁷⁸

If the argument I have presented here is right, Dokic and Engel have failed in vindicating RP. I will discuss now an alternative strategy for defending RP, proposed by Whyte (1990).

⁷⁷ This is a rather trivial claim: since all those implicit beliefs are derived from the success of the action, it is to expect that, when such implicit beliefs are present, the action is successful.

⁷⁸ Maybe it could be suggested that when there are ignored obstacles, the action is not properly derived from the agent’s beliefs (these beliefs do not offer proper grounds for the performance of the action). Nevertheless, Dokic and Engel do not argue in this direction.

5.4.2.2 *No-obstacles auxiliary belief*

One alternative way of tackling the problem of ignored obstacles is to resort to some general 'no-obstacles' belief, that is, a belief stating that there are no impediments for the success of the action – as long as the rest of the beliefs the action follows from are true, of course. This is what Whyte (1990, 1997; also Mellor 2003) proposes. Brandom (1994b) acknowledges that the introduction of such a belief would eliminate the problem of ignorance, but at the risk of making the principle uninformative – Brandom claims that introducing a no-obstacles belief 'trivializes' RP (Brandom 1994b: 177). Brandom's criticism goes along the following lines. It seems, on the one hand, that one can reformulate the no-obstacles belief as the belief that my action will be successful provided that the other beliefs it is based on are true. Then, for any set of beliefs $\{B_i\}$ and any action A_j , RP plus the no-obstacles belief, could be expressed as: "If $\{B_i\}$ is true, and the belief that (A_j will be successful when $\{B_i\}$ is true) is also true, then A_j will be successful."⁷⁹ Now, the principle just codifies a good deductive inference:

(1)

$\{B_i\}$

$\{B_i\} \rightarrow A_j$ succeeds

A_j succeeds

⁷⁹ It may be argued that the no-obstacles belief does not appeal to the truth of my other beliefs, but rather to the facts they express. So, it would be a belief that my action will be successful – given that $p, q...t$ (where $p, q...t$ are the contents of the other beliefs from which my action is derives; or, if preferred, they are what such beliefs represent). However, I am assuming that if – and only if – the belief that p is true, then p is the case (this is quite uncontroversial). In this way, both versions of the no obstacles belief would be equivalent (Whyte (1997) misreads Brandom's criticism (1994b) in this point).

(1) will be good independently of the content of the beliefs $\{B_i\}$ or the action A_j , so, in this sense, it is uninformative and does not add anything to our knowledge of the truth conditions of the beliefs involved (one could connect through this principle any random beliefs and actions: for instance, the belief that today is sunny and my success or failure in winning the lottery).

Brandom is certainly right in claiming that the validity of (1) does not depend on the contents of $\{B_i\}$ or on what action A_j is. But concluding from this that inferential schemas such as (1) cannot be of use for determining the truth conditions of the beliefs $\{B_i\}$ is too hasty.⁸⁰ After all, it seems that one can isolate the truth conditions of $\{B_i\}$ – factoring out the contribution of the conditional belief – by the methods of detachment described in section 5.4.1.

More specifically, the strategy for detaching the truth conditions of the non-conditional beliefs $\{B_i\}$ idea would be to look at further instantiations of RP in which $\{B_i\}$ recommend the performance of *different* actions.⁸¹ There will still be no-obstacles conditional premises involved, but these will be different from the one present in the original inference. So, for instance, the truth of the belief “The water in the pot is boiling” plus the truth of the conditional premise “If I put pasta in a pot and the water in the pot is

⁸⁰ Assume that the conditional premise “ $\{B_i\} \rightarrow A_j$ succeeds” were true in all circumstances. Then, the truth conditions of the conclusion “ A_j succeeds” would offer non-trivial information about the truth conditions of $\{B_i\}$ – in particular, they would inform us about those circumstances in which $\{B_i\}$ are false, namely those circumstances in which A_j is not successful. Analogously, the truth conditions of “London is to the North of Madrid” provide information about the truth conditions of “Madrid is to the South of London” despite (2) below being formally good, in the same sense that (1) is:

(2)

Madrid is to the South of London

If Madrid is to the South of London, then London is to the North of Madrid

London is to the North of Madrid

⁸¹ This strategy was already introduced in section 5.4.1, when dealing with ‘stability conditions’, which are just no-obstacle conditions.

boiling then the pasta will get cooked” guarantee the success of my action of putting the pasta in the pot so as to cook it. But entitlement to the belief “The water in the pot is boiling” will also give reasons to put a thermometer in the pot, if I want to calibrate the thermometer by fixing the 100°C measure. The conditional premise here would be something like “If I put the thermometer in boiling water, it will show the measure corresponding to 100°C”. So, the non-conditional premise would be the same as in the previous inference (“The water in the pot is boiling”) but the conditional premise would change. It seems that, by comparing the truth conditions of several such sets, one may be able to single out the truth conditions of the non-conditional premise “The water in the pot is boiling.”

The idea is that these actions have *different* obstacles, so that the corresponding conditional premises are falsified in different conditions. For instance, salt in the water will slightly raise the boiling temperature – so that a thermometer will not be successfully calibrated at 100°C in salty water (at least, if you want the calibration to be very precise). In contrast, boiling salty water may be perfectly suitable for cooking pasta. On the other hand, a substance which did not affect the boiling temperature but made pasta inedible would be an obstacle for cooking pasta, but not for the calibration of thermometers. One will be able to single out the truth conditions of the common non-conditional premises when the conditions in which each of the conditional premises is falsified are complementary – that is, when their disjunction is not falsified in any conditions.⁸²

Brandom does not consider this possibility, but he should, since this is the kind of method of detachment that he endorses. He explicitly says that the truth conditions of an individual belief p may be isolated ‘by considering conjunctions that differ from a given one only by dropping p , or by substituting some other belief for it’ (1994b, 175; my emphasis). This is precisely what one would do in the approach presented in the paragraph above: one considers conjunctions that differ only in that the conditional no-

⁸² This follows from a general point about the truth conditions of conjunction that I made in section 5.4.1. If “[$B1 \rightarrow A$] or [$B1 \rightarrow A'$]” is true in all circumstances, and one knows has the accuracy conditions of “[$B1$ and [$B1 \rightarrow A$]]”, and of “[$B1$ and [$B1 \rightarrow A'$]]”, then the the accuracy conditions of $B1$ can be found out by looking at the disjunction of both conjuncts: “[$B1$ and [$B1 \rightarrow A$] or [$B1 \rightarrow A'$]]” will have the same accuracy conditions as “ $B1$ ”.

obstacles premise is substituted by another one. However, when discussing no-obstacles beliefs, Brandom says:

one cannot, instead of dropping it [the no impediments belief], substitute some other belief for it and observe the difference in what conditions must hold for the resulting conjunction to guarantee the truth of actions based on it. For any other belief, put in its stead, will not rule out the possibility of failure due to ignorance rather than false belief

(Brandom 1994b: 178)

Well, of course the substitution of a different no-impediments (i.e. no-obstacles) belief will not guarantee the success of the action recommended previously; but it may guarantee the success of some new action. So, if the no-obstacles conditional premise “If I put the pasta in a pot with boiling water, the pasta will get cooked” is substituted by “If I put the thermometer in boiling water, it will show the measure corresponding to 100°C”, this new conditional premise, plus the premise “The water in the pot is boiling” will guarantee that the thermometer can be calibrated at 100°C by putting it in the water – even if it will stop being guaranteed that the pasta will get cooked if put in the pot.

This sort of detachment strategy, however, faces serious difficulties – arising from the possibility of an indefinitely large number of possible obstacles for any given action. Above I said that the way of detaching the truth conditions of a non-conditional belief B1 was to look at the truth conditions of its conjunctions with conditional beliefs falsified in different circumstances – ideally, complementary beliefs, so that at least one of them is accurate in all circumstances. The idea is the following. Think of a possible obstacle for the success of an action derived from a given belief; for instance, the action of cooking pasta in boiling water will fail if the water contains some substance that does not alter its temperature, but prevents the pasta from cooking. In order to screen out the effects of this possible obstacle, one would consider other actions whose success is compatible with

the presence of such an obstacle.⁸³ For instance, even if the anti-cooking substance is in the water, one could still measure 100°C by introducing a thermometer in the pot.

If there were a definite number of possible obstacles for each action, this strategy might work. The problem is that there will always be further possible obstacles for the two actions compared and, in particular, nothing excludes the possibility that there is a common obstacle for both actions. For instance, extremely low pressure will make the temperature of the boiling water drop significantly. This would be an obstacle for both the action of cooking pasta and the action of measuring 100°C with the thermometer. Of course, there are other actions derived from the belief that the water is boiling whose success is compatible with extremely low pressures. But there may be further common obstacles. So, no matter how many alternative actions are considered, it is always possible that there are some further defeating conditions which have not been factored out yet – and which will, thus, muddle the assignation of truth conditions to the belief we are interested in. One would never finish detaching the conditions of failure associated with all the different possible obstacles for each action.

So, the prospects of factoring out the truth conditions of Whyte's no-obstacles belief are bleak – even if Brandom's rejection of Whyte's proposal is too quick. While the detachment methods reviewed in section 5.4.2 can be employed so as to factor out the contribution of no-obstacle beliefs (*pace* Brandom 1994b), one would never complete such process of detachment, because of the indefinitely large number of possible obstacles for (I am assuming) most actions.

This is a general problem for success semantics, though, rather than just a weakness peculiar to Whyte's proposal. Any attempt to detach background anti-obstacle conditions for the success of actions will face the same sort of difficulty: the large number

⁸³ To be precise, the effect of a certain obstacle will only be completely screened out if one considers alternative actions which are properly derived from the belief in question in *all* circumstances in which the said obstacle is present. This does not happen with the example of the thermometer. The action of measuring 100°C will fail, for instance, if the water contains the anti-cooking substance and there is some other substance that alters its boiling temperature. We would have to look for further actions in order to deal with these circumstances. (As suggested in note 9, another possibility is to accept Lewis' discrimination of a privileged class of 'natural' properties).

of possible obstacles (and, therefore, of required anti-obstacle background conditions) prevent the process of detachment from having a definite end. This will be so even if it were the case that agents have beliefs – implicit or explicit – about all the conditions necessary for the success of the relevant action – in particular, beliefs about the absence of each of the possible obstacles for such an action (e.g. implicit beliefs à la Dokic and Engel). Granting that there is no end to the possible obstacles that may come up, there will be no end to the background conditions for the success of a given action, either. Thus, having beliefs about all these background anti-obstacle conditions would amount to having an indefinitely large number of collateral beliefs. There would be no definite end to the detachment of the truth conditions of such collateral beliefs by means of the methods reviewed in section 5.4.1. Imagine that we want to factor out the background effect of gravitational forces; in order to do so, we imagine – or perform – a practical experiment in a spaceship. But this alternative context would not manage to get freed of other background conditions – such as the absence of mischievous magic spells. We may suggest further alternative circumstances, but – under the present assumptions – there is no certainty that we have ever detached every possible background condition.

Thus, the problem of ignored obstacles does not vanish if it is granted that the agent has some sort of implicit beliefs about the background conditions for the success of his action: the truth conditions of such implicit beliefs would still merge with the truth conditions of the explicit beliefs of the agent, and there seems to be no easy way of completely detaching them. The sort of holism imposed by success semantics may be more radical than what was expected.

It seems, therefore, that RP cannot yield truth conditions for individual beliefs. In the section 5.5 I will discuss how Simon Blackburn has modified that principle so as to immunize it against the problem of ignored obstacles.⁸⁴ I will argue that Blackburn's

⁸⁴ Of course, one may try to keep looking for alternative defenses of the thesis that truth guarantees success – maybe adding certain provisos. For instance, Nanay (2013) has suggested that a possibility is to say that, at least for some kinds of mental representations, truth raises the probability of success – rather than guarantees it. I will not discuss this proposal in detail, but I think that it is not promising (in the presence of an ignored obstacle, some false representation may offer a higher probability of practical success than an accurate representation).

approach is also problematic – and, in particular, falls prey to similar difficulties to those faced by Whyte’s proposal. Before doing so, I will briefly introduce a further problem for Whyte’s account – a problem which will reappear in Blackburn’s approach.

5.5 Truth as a necessary condition for success: revisiting the question

In section 5.3. I argued that practical success may be the result of lucky accidents, even if the agent’s beliefs are false (e.g. I falsely believe that the London train is due to set off at 4 p.m., when it is actually due at 3 p.m.; luckily for me, it is accidentally delayed and I am able to catch it at 4 p.m.): thus, truth is not necessary for the success of an action.

In standard versions of RP (for instance Dokic and Engel (2002)), the strategy for identifying success based on falseness was to note that RP guarantees success to all the actions that are derived from a given set of true beliefs – not just to some of them: if a belief is false, some of the actions derived from it will most likely fail (even if some other of the derived actions may be successful despite the falseness of the belief). However it is doubtful that this strategy is available to Whyte, once he introduces general no-obstacles belief. After such auxiliary beliefs are introduced, it is not clear that one is in a position to discriminate easily cases of success based on false beliefs. From the conjunction of a set of beliefs $\{B_i\}$ and the corresponding no-obstacles auxiliary belief (which I have been representing as: $\{\{B_i\} \rightarrow A_j \text{ succeeds}\}$) only one action follows, namely A_j . Thus, we cannot appeal to further actions derived from such a conjunction of beliefs (including the no-obstacle belief) in order to check whether success in performing A_j was based on false beliefs. Sure enough, the set $\{B_i\}$ will usually be involved in the performance of further actions – call them A_k . But, in Whyte’s proposal, these further actions will follow from $\{B_i\}$ plus a new no obstacles belief $\{\{B_i\} \rightarrow A_k \text{ succeeds}\}$ – a no obstacles belief that differs from the one relevant for the derivation of the former action, A_j . Now, if A_k fails, it can be because $\{B_i\}$ was after all false, or because the new no obstacles belief $\{\{B_i\} \rightarrow A_k$

An alternative approach is to try to relativise Ramsey's Principle to some 'normal' context of application – in which no unforeseen ignored acts appear. See Hattiangadi (2007) and Loewer (1997) for an account of the difficulties faced by this kind of projects.

succeeds) was false. Thus, a failure in some further action derived from $\{B_i\}$ will not let us know whether the success of the first action A_j (derived from $\{B_i\}$ and the corresponding no obstacles belief) was based on false beliefs.

A version of this problem affects Blackburn's (2010: 187) reformulation of success semantics, which I will present in the next section.

5.6 Truth without guarantee of success

In view of the problem of ignored obstacles, it can be accepted that truth is neither necessary nor sufficient for success. Is anything useful still left in success semantics if we adopt this defeatist perspective? Simon Blackburn (2010: 181-199) thinks there is. He renounces the attempt to explain practical failure by appeal to false beliefs – since it can also be caused by ignored facts. Instead, Blackburn focuses on cases of success, and argues that the state of things represented by a belief is some state of things that typically figures in the explanations of the success of the actions based on such a belief. Blackburn prefers to go down to the sub-sentential level,⁸⁵ where he gets at the following revised version of RP (which he calls the 'fundamental schema'):

Suppose the presence of 'a' is a feature of a vehicle 'a...'. Then 'a' refers to a if and only if actual and possible actions based upon the vehicle 'a...' are typically successful, when they are, at least partly because of something about a.

(Blackburn 2010: 187)

⁸⁵ Blackburn (2010: 186-187) prefers to formulate the fundamental schema at the sub-sentential level so it is easier to deal with seldom (or never) used sentences.

5.6.1 Success despite falseness

The modesty of Blackburn's schema may reintroduce some problems, though. As I pointed out above, some successful actions will be based on false beliefs – maybe we just get lucky and the world helps us to succeed despite our false beliefs. The success of an action based on a false belief that p will not be explained because of something about p (the belief is false, so p does not take place). In Blackburn's sub-sentential language: when a , the intended referent of 'a' is not present, the success of an action based on a 'vehicle' which includes an occurrence of 'a' will not be explained appealing to something about a .⁸⁶ For instance, imagine that I believe that the flask is full of orange juice, so I try to quench my thirst by drinking its contents. Assume that my belief was false, there was no orange juice in the flask. However, there was water, so I succeeded in quenching my thirst. This success had nothing to do with the presence of orange juice, the referent of 'orange juice' in the belief carried by the vehicle 'The flask is full of orange juice'.

When I discussed this issue in section 5.3, the way to tackle the problem was to note that, if a belief is true, *all* the actions that follow from it must succeed – while *some* of the actions based on a false belief will fail, even if not necessarily all of them. However, Blackburn cannot resort to this strategy, since he does not defend the view that practical failure is always due to false beliefs. Thus, Blackburn has to admit that both false and true beliefs may have successful and failed practical consequences.

How are we to distinguish, then, the cases of success that determine the referent of a belief, from cases of success despite the falsity of our beliefs (despite the absence of the intended referent)? Sure enough, Blackburn admits that the patterns of explanation described by his schema will only apply 'typically'. But this hardly helps, for he still has to specify which situations are typical. If an unqualified appeal to typical cases were acceptable, we could as well say that truth guarantees success in typical (or standard) cases – in order to get away from the problem of ignored obstacles.

⁸⁶ Of course, I can have beliefs about the *absence of a*, which will lead to successful actions precisely when a is absent. For the sake of simplicity I will focus on beliefs whose accuracy depends on the presence of what is represented by 'a'.

Moreover, practical success based on false beliefs is not always accidental or random: in some cases it can be quite pervasive and systematic, so typical success cannot be directly equated with statistically frequent success.⁸⁷ Imagine that there are two kitchen pots in a shop, made of two different materials, material A and material B. Even if their microscopic structure is very different, both materials share a lot of macroscopic properties. In particular, they share most of the properties that affect my practical dealings with them in the kitchen: they have the same weight and size, they are scratched by the same objects, both are good heat conductors. The only difference relevant in a kitchen scenario is that material A is suitable for use in microwave ovens, whereas material B pots melt in such ovens. Suppose now that I mistakenly believe that I bought pot A, when as a matter of fact I bought pot B. Most of the actions based on my false belief that I am using pot A will be successful – at least in a kitchen environment. The pot will be able to contain as much food as I thought; it will not be scratched by my metal fork, and it will be heated up in the fire as I expected. The only action that will fail will be to use it in the microwave oven (something I rarely do). But Blackburn cannot say that my belief was false just because it led to a single failed action (in subsentential terms: Blackburn cannot say that the vehicle ‘pot A’ in my belief did not refer to the pot in my kitchen just because one action based upon such a vehicle failed). Therefore, it seems compatible with Blackburn's framework to conclude that my belief was true, and that what is represented by it actually took place. Maybe Blackburn would say that the vehicle ‘pot A’ figuring in my belief “I am using the pot A” actually represented the pot of material B, since it was this pot that explained the success of most of my actions – and it seems legitimate to call ‘typical’ such pervasive and systematic patterns of success. Notice, however, that it is implausible to attribute to me a belief representing pot B, since I was disposed to use the pot in the microwave (when I am perfectly aware that pot B is not suitable for this). As for the failure of the microwave oven action, Blackburn would have nothing to say, since his schema only deals with successful actions: failure may have many different origins.

⁸⁷ See Millikan (1989) for an analogous point in relation to proper and statistically frequent ways in which a trait may function.

Imagine now that I have both pots, and that roughly half of the time I believe I am using pot A I am right, and half of the time wrong – it is actually pot B. It is not clear what ‘typical’ success would be in this example. The number of successful actions following from my true beliefs in pot A could be, more or less, the same than the number of successes resultant from my false beliefs – that is, from my beliefs of being using pot A when it is actually pot B. Actually, if I happen to be mistaken more often than not, the existence of pot B would explain most of the successes of my beliefs about being using pot A – and, consequently, that would be what my beliefs are about.⁸⁸ If the difference between typical and atypical success is related to the frequency of the success, then highly contingent matters, such as the number of times I happen to find one or the other pot in the cupboard, would determine what is represented by certain kind of belief – and, anyway, we would have to fix arbitrarily some threshold frequency: maybe 50% of successes is enough to be typical, or maybe we prefer to fix the demarcation in 75%, or 90%.⁸⁹

It might be replied that the agent in this example actually has beliefs about the macroscopic properties shared by material A and B (or something like that); but, again, this does not seem right. The fact that the agent is thinking of a pot made of material A is revealed by her practical disposition to use the pot in the microwave oven. If her belief were about a pot of material B, or about the properties shared by both materials, the agent would not think that the pot was suited for microwave cooking – so she would not have tried to cook in the microwave with it. If this distinction exists, it is implausible to

⁸⁸ One may want to complain that this example only focuses on the actions of one agent – and that, overall, the success of the actions of most people thinking they are using pot A will be because they are in fact using pot A. But one could also think of cases in which most relevant agents are mistaken about something but, despite this, their ensuing actions have a high success rate. Imagine a situation in which most scientists accept the existence of some entity or mechanisms (say, phlogiston or supersymmetrical particles), and that this acceptance leads to many successful actions (experimental manipulations), and only few failures. It could be, nevertheless, that such entity or mechanism turned out not to exist after all – and that the scientist’s practical success was explained by some alternative entity or mechanism.

⁸⁹ Millikan’s teleosemantics (1984) is a similar proposal to Blackburn’s. Millikan, however, can specify a non-statistical notion of typicality (what she calls ‘Normal’ conditions) in relation to the functioning of some biological device, by focusing on the conditions that took place in cases of success that led to the selection of the device.

maintain that, when thinking about material A, agents are really thinking about the properties shared by A and B.

Remember that these problems do not affect the strong versions of success semantics encoded by RP, according to which the truth of our beliefs is incompatible with practical failure.⁹⁰ In that case, a single failure derived from my belief of being using pot A would mean that my belief was wrong – even if it lead to many instances of success. In section 5.4.3. I noted, however, that similar difficulties are faced by Whyte’s sophisticated formulation of RP – in which no-obstacles beliefs are introduced.

5.6.2 Ignored obstacles again

It is not surprising that Blackburn’s and Whyte’s proposals face similar difficulties, since both approaches are actually closer than what one might think at first. In this way, Blackburn’s proposal ultimately shares the problems found by Whyte’s (and also Dokic and Engel’s) version of RP in disentangling background anti-obstacle conditions from the success conditions of actions. Let me show why this is so. Blackburn claims that (typically), if an action derived from a vehicle ‘a...’ is successful, what is represented by such a vehicle will actually obtain (more specifically, the referent of that vehicle will be involved in bringing about the success of the action, at least in the case of affirmative beliefs). Blackburn may well be right about this much; but even if this is so, it does not mean that one is in a position to identify the referents of the relevant vehicle by looking at the conditions of (typical) success of the derived actions. Typical cases of success are explained because of the presence of the referent of the vehicle *plus the occurrence of the relevant anti- obstacle conditions*. How are we to disentangle the referents of the vehicle from these background no obstacle conditions? This is a version of the problem that arose in relation to Whyte’s reformulation of RP: Whyte argues that – barring cases of success despite falsity – actions succeed when they follow from a set $\{B_i\}$ of true beliefs and, moreover, there are no obstacles for such actions (i.e. the corresponding no

⁹⁰ Of course, as I have explained, these strong versions are vulnerable to the problem of ignored obstacles. What I want to point out is that Blackburn, trying to avoid that problem, falls into other difficulties.

obstacles belief is also true). So, the problem already faced by Whyte's reformulation of RP resurfaces in relation to Blackburn's proposal: it does not seem possible to disentangle the referent of a vehicle 'a...' from the background anti-obstacle conditions required for the success of the actions in whose derivation 'a...' is involved (at least if the number of possible obstacles is granted to be in principle indefinitely large).

Blackburn could try to say that referent of the vehicle 'a...' is what is invariant across the cases of (typical) success of all the actions following from the vehicle (whereas anti-obstacle conditions would change from action to action). This would be analogous to Whyte saying that the truth conditions of $\{B_i\}$ can be isolated by contrasting the conditions of success of *all* the different actions following from that set. But, as we saw above, no matter how many different actions one contrasts, there is no guarantee that there remain common obstacles to all those actions – and thereby, common background anti-obstacle conditions that have not been screened out. This is to say, it could always be that the conditions invariant across cases of success of (no matter how many) different actions following from the vehicle 'a...' are not identical to the referent of 'a...', but still include some further background anti-obstacle condition (say, that oxygen is present in the surroundings, that the temperature is not extremely low and so on and so forth).

Thus, Blackburn's approach is also affected by the general problem undermining success semantics – discussed at the end of section 5.6.⁹¹ If Blackburn's proposal were able to overcome such a problem, so would Whyte's reformulated version of RP. A comparative advantage of Blackburn's proposal is that it does not require agents to have general beliefs about the absence of obstacles; but, if one's views about what it takes to have such no-obstacles beliefs are not too demanding, this requirement will not be so cumbersome after all. On the other hand, both Blackburn's and Whyte's proposals have

⁹¹ Actually, a similar version of the problem comes up as well in the case of Millikan's teleosemantics (1984). Millikan argues that the accuracy-conditions of representations are those conditions that explained the success of the relevant actions in those cases which contributed to the selection of the mechanisms producing such actions. But these conditions will include indefinitely many background anti-obstacle conditions which, intuitively, one would not want to mingle with the accuracy-conditions of the representation (e.g. conditions concerning the existence of gravitational forces, oxygen in the atmosphere etc).

to show that they are able to discriminate cases of success based on false beliefs – and it seems doubtful that they have resources to answer this challenge.

To conclude, the possibility of indefinitely many obstacles for the success of most actions poses serious difficulties for the project of deriving the truth conditions of beliefs from the conditions of success of actions following from them: those conditions of success will be partly constituted by background anti-obstacle conditions, which cannot be easily factored out. This is a problem that no available version of success semantics seems to be able to address properly.

In the following section I will turn to a further problem faced by success semantics, namely the determination of the conditions of success of actions.

5.7 Conditions of success

Let us grant that the conditions of success of our actions manage to provide truth conditions for the beliefs they are based on. Even if this is the case, the success semanticist will still have to explain what makes an action successful. For the sake of simplicity, think of practical success as the satisfaction of the desires that motivated the action. Conditions of truth, therefore, would be accounted for in terms of conditions of satisfaction of desires. However, the concept of satisfaction of desires is itself clearly a representational semantic concept,⁹² so success semantics would not be offering a reductive explanation of representation unless it offers a suitable reductive account of the conditions of satisfaction of desires. Of course, a modest success semanticist may only aspire to illuminate the relations between different semantic concepts (conditions of truth of beliefs and conditions of satisfaction of desires), but this would severely limit the scope (and interest) of the theory.

This pressing question has been forcefully exposed by David Papineau: 'But the notion of satisfaction, for desires, is closely analogous to that of truth, for beliefs. And, in

⁹² A desire can be viewed as representing a state of affairs that would satisfy it. The conditions of satisfaction of the desire would precisely specify which this state of affairs represented by the desire is.

particular, it is a notion that raises just the same conceptual problems about representation' (1987: 56). Most proponents of success semantics acknowledge this difficulty, and try to suggest ways to overcome it.

5.7.1 Reductionism

The most immediate solution for the problem I have outlined is to find a straightforward reductive account of the satisfaction of desires, or at least of some 'basic desires' (Whyte 1991). One simple proposal is to say that a desire is satisfied when it becomes extinct. So, the fact that eating an apple makes my craving stop, would imply that I had a desire to eat an apple. Nevertheless, this approach faces obvious objections, already presented by Wittgenstein (see Dokic and Engel 2002): in general, my desire can disappear without the circumstances that would have satisfied it taking place. My desire to eat an apple may vanish after eating a big salad – or after going to sleep. I may get bored of waiting and change my mind. It seems that these problems affect all kind of desires, even the most 'basic' ones – although I will not argue for such a claim here.

Whyte (1991) tries to deflect this kind of objections by distinguishing different ways in which a desire may disappear. When a desire gets extinct through its proper fulfillment, the behavior of the agent will be reinforced: my disposition to act in similar ways when I have such a desire will be enhanced. In contrast, even if a desire to eat an apple may go away with a kick in the stomach, my disposition to act in such a way will not be reinforced. However, there seem to be ways of extinguishing desires without fulfilling them which can result in a reinforcement of the behavior that made the desire stop: imagine that my desire of chocolate disappears when taking some (harmless) drug. My disposition to respond in this manner to chocolate desires may be reinforced: perhaps I think that eating chocolate is not healthy, so I prefer to silence my desire by means of the drug.

5.7.2 Holism

If one thinks that direct reductionist accounts of the conditions satisfaction of desires are not promising, an alternative is to look for a holist way of determining such conditions (of course, without appealing, ultimately, to representational notions). Even if the conditions of satisfaction of individual desires cannot be fixed in isolation, maybe there is some way of fixing the conditions of satisfaction of systems of desires.

The model here, of course, would be Davidson's theory of charitable interpretation, according to which the content attributed to the beliefs of an agent by an interpreter would be such that the truth of those beliefs is maximized. We would interpret people in a way that makes them as rational and truthful as possible, in light of the evidence available: most sentences would be true when a speaker takes them to be true. A similar strategy might work for desires (Dokic and Engel 2002): we attribute desires to agents so that such desires are typically satisfied when agents take them to be – in particular, when they take them to be extinct. Maybe, in some specific instances, a desire is extinct but not properly satisfied – or the other way round –, but on the whole, these cases would be minimized by our interpretative choices.⁹³

This strategy is explicitly proposed by Dokic and Engel (2002), and also suggested by Blackburn (2010: 194-195). Blackburn describes a toy model of linguistic communication, in which adults would develop interpretations of the behavior and speech of a baby. At first, it is quite indeterminate what the baby wants when she utters some sounds. However, after some time, we will observe that the baby tends to be pacified by certain things after uttering such sounds, and stable patterns of interpretations will begin to emerge.

The problem with this proposal is that, if one adopts a Davidsonian view in relation to the content of desires, it is not clear why one should not just endorse the Davidsonian account of representation all along. It would seem that the resulting version of success

⁹³ 'It may be a normative principle of a pragmatist theory of interpretation that most "extinct" desires are really satisfied, even if one does not know which ones are' (Dokic and Engel 2002: 63).

semantics is not very different from a Davidsonian position.⁹⁴ Assessing the merits and problems of the Davidsonian program is beyond the scope of this dissertation; for my purposes here, it is enough to say that by embracing such a Davidsonian outlook, success semantics would stop offering a distinctive account of representation.

Of course, one can just take success semantics to impose a constraint on our attributions of content to beliefs and desires. The contents of desires and beliefs would not be independent from the content of other beliefs and desires. Even if such constraints can illuminate in interesting ways many features of our discursive practices, most likely this result will be regarded as a retreat by those who attempted resort to success semantics in order to give a reductive account of representational content – that is, content which is made right or wrong (true or false) by the way the world is.

5.8 Conclusions

In this chapter I have presented some of the most relevant versions of success semantics – that is, theories trying to account for representational content in terms of practical success – and I have examined some of the problems that such theories face. Of course, there are more difficulties than the ones I have considered here. Just to mention a few additional worries, some beliefs may not have direct practical implications, and there is also the question of beliefs about the past or the future.⁹⁵ However, I have discussed what I take to be two of the most pressing problems for success semantics: ignored facts and the specification of the conditions of success of actions. Both problems require a convincing answer from success semanticist – something that has not always happened.

In relation to the first problem – that ignored facts may thwart our actions despite even if they are based on true beliefs –, I have tried to show that the solutions proposed

⁹⁴ Dokic and Engel (2002) seem to accept this integration of success semantics in Davidson's framework. They claim that success semantics helps to develop and illuminate Davidson's interpretative program, since it is often easier to perceive in agents the attitude of taking a desire to be satisfied than the attitude of accepting the truth of a belief.

⁹⁵ I think that these are important – though tractable – issues, but I will have to leave them for another occasion.

by Dokic and Engel (2002), Whyte (1990, 1997) and by Blackburn (2010) are not satisfactory. In particular, there are seemingly insurmountable difficulties in detaching the truth conditions of beliefs from background no-obstacles conditions, when deriving truth conditions from success-conditions. I will present my own approach to this issue in chapter 7.

The second pressing problem for success semantics discussed here, is that it puts too much weight in the notion of success of an action, without offering a suitable account of what the conditions of success of an action are. The notion of conditions of success – or, if we prefer, conditions of satisfaction of our goals – seems to be as representational as the notion of conditions of truth for beliefs. Therefore, it does not constitute a safe resting point for a reductive explanation of representation.

Insofar as success semantics intends to find such reductive account of representation (Whyte 1990), it would seem that it must give a reductive analysis of the conditions of success of actions. I have argued that the simple reductive analysis proposed by Whyte (1991) – according to whom conditions of success should be identified with conditions of extinction of our motivating desires – is unsatisfactory.

Finally, I have briefly presented holist analysis of the conditions of success of actions – as found both in Dokic and Engel (2002) and Blackburn (2010). It is hard to evaluate this sort of holist approach unless a more detailed and developed proposal is at hand. Again, such a project would have to be undertaken elsewhere.

An alternative is to turn to evolutionary accounts of the goals of biological systems, and in particular of the conditions of satisfaction of desires (Papineaus 1987; Millikan 1989). I will examine these proposals in the next chapter. I will discuss the shortcomings of such approaches and I will how indicate how overcoming these shortcomings motivate my own account of goals – which I will put forward in chapter 8.

Success semantics, thus, has to face what seem to be formidable obstacles. This does not mean – I think – that this theory, or in general pragmatist accounts of representation, do not merit further investigation. On the contrary, I still regard them as the most promising roots for a satisfactory (and science-compatible) account of the way human beings make representations of their environment, as part of their interaction with the world. In the chapters 7 and 8, I will try to elaborate a pragmatist account of

representation which deals with the obstacles encountered by success semantics and other related approaches.

Chapter 6: Biological Goals

6.1 Introduction

In chapter 5 I discussed the difficulties faced by success semantics in relation to the determination of the conditions of success of actions. Success semantics tries to account for the representational purport of beliefs by specifying their truth conditions in terms of the conditions of success of the actions that follow from them. The problem, pointed out by Papineau (1986) and others, is that the notion of conditions of success seems to be as representational as the notion of truth conditions.

One possible way of dealing with this issue is to try to relate practical success to the achievement of some biological goal. For instance, Papineau suggests that 'we can think of the satisfaction condition of a desire as the effect it is biologically supposed to produce' (1987: 67). The idea, therefore, is to understand the truth conditions of beliefs in terms of the success conditions of the actions derived from them – in line with success semantics – , and to account in turn for success conditions in terms of the achievement of biological goals.

In this chapter I will examine this way of addressing the problem of determining success-conditions. On the views of biological goals I will consider, goals are taken to be set either by natural selection (Millikan 1984; Neander 1991) or by the process of self-organization and self-maintenance of organisms (Christensen and Bickhard 2002; Weber and Varela 2002; Di Paolo 2005; Barandarian and Egbert 2014; Mossio et al. 2009). I will argue that the account of practical success provided by these approaches is unsatisfactory from the sort of pragmatist perspective I am assuming. My main argument will be that it is not always clear that the sort of biological goals described by these approaches are goals attributable to *individual* agents and that, furthermore, the satisfaction of such biological goals is too remote from the behavior and attitudes of the agents subject to those goals.

The inadequacies of these accounts of biological goals – at least, when they are seen as grounding intentionality – will serve as motivation for my own proposal, which will be presented in chapter 8.

In section 6.2 I motivate the project of appealing to biological goals in order to provide an account of intentionality. In section 6.3 I present the two views of biological goals that I will consider (evolutionary and organizational accounts). After that, in section 6.4 I argue that evolutionary views do not offer a good basis for developing an account of individual intentional agency, since many of the goals they identify should be attributed to lineages or reproductive chains rather than to individuals. In section 6.5 I claim that related problems affect organizational accounts.

6.2 Practical success as the fulfilment of biological goals

Biologists typically talk of goals when describing biological systems and their behavior: teleological vocabulary is widespread in biological explanations. For instance, it would not be unnatural to say that some protozoon swims around with the goal of finding food, or that a zebra runs with the goal of escaping from a predator. The goals ascribed to biological systems involve, in many cases, references to features of the environment – e.g., nutrients to be ingested, threats to be avoided or shelters to be found. The achievement of such goals depends on how things stand in the world – on whether nutrients are actually ingested or shelter is found. Consequently, biological goals seem to be a good place to look at in order to characterize a minimal notion of intentionality – that is, of behaviors or states of a system which are *directed towards* features of the environment.

In order to get from biological goals to the expression of intentional content, one may appeal to some version of success semantics. According to success semantics, what semantic content (more specifically, truth-conditions) is expressed by some belief is determined by the success-conditions of the actions following from that belief. The problem discussed in 5.7 was that it is not obvious how to identify the success conditions of the actions following from a given belief, at least in a way that does not presuppose that the belief has certain content. If it is possible to appeal to some notion of biological goal, however, it seems that there is a solution to this problem: the success conditions of actions can be identified with the conditions of achievement of the biological goals associated with the performance of those actions. Since biological goals, as I mentioned

above, depend on the way things are in the environment, success conditions specified in this way would seem to incorporate the required friction with the world. The success of actions would depend on how the world is, and in turn success conditions would determine the truth conditions of the beliefs leading to those actions. The truth of such beliefs, therefore, seems to depend on the way the world actually is, as needed if the objectivity of representational practices is to be respected.

Actually, teleosemantic accounts of representation of the sort Millikan develops (1984, 1989a, 1989b, 2005) can be seen as versions of success semantics in which the conditions of success of actions are settled by the biological functions of the mechanisms performing those actions. So, the truth conditions of a representation would be given by the conditions in which the systems using that representation manage to succeed in performing their biological functions. In particular, teleosemantics would offer a reductionist account of the satisfaction of desires.

Millikan's theory of representation and success semantics, thus, share the pragmatist insight that both the status of something as a representation, and its content, derives from its uses. As Millikan puts it, in order to characterize representation and representational content, one should pay attention to 'representation consumption, rather than representation production' (1989: 283). Millikan distinguishes a system producing the representation (producer) and a system using the representation for some purpose (consumer system). The central idea of Millikan's proposal is that accurate representations guide consumer systems in successfully performing their functions. So, the conditions of accuracy for a representation are those conditions that need to obtain in order for the consumer system to perform successfully its functions – more precisely, those conditions that explain the successful functioning of the consumer system in normal cases, that is, excluding accidental or random success. This is very similar to Blackburn's version of success semantics (discussed in 5.6): the main difference is that Millikan appeals to the notion of biological function in order to characterize success conditions, while Blackburn does not.⁹⁶

⁹⁶ Appealing to biological functions comes with a bonus: Millikan can characterize typical cases of success as those cases in which success is achieved in virtue of the relevant biological mechanisms working as they are supposed to work (that is, by performing their biological functions, rather than just by luck or accidentally). Remember that in section 5.6 I argued that characterizing the notion of typical success was problematic for Blackburn.

On the face of it, the combination of success semantics with a specification of success conditions in terms of biological goals seems promising. However, this approach is not unproblematic. First, there is the worry that many intentional actions (say, switching on your computer) do not have obvious connections with biological goals. Although I agree that this is a problem, I will not press it: I will grant that it is somehow possible to start with an account focused on basic actions more clearly relevant for the biological goals of agents, and from there get to a sophisticated view addressing more complex behaviour. Millikan (1984; 2005) has tried to sketch such a sophisticated account. As she insists (1984: 45-49; 1989), many biological systems are designed to learn and to adapt to new situations. According to her, the resulting behaviours, even if novel, may have functions derived (more or less distantly) from some biological function.⁹⁷ As I said, I will grant that a project along such lines is on the right track.

The problem I will focus on, instead, is that it is not clear how to characterize biological goals (in a way that makes it possible to construct an account of intentional agency in terms of such goals). Biological systems are (as far as we know) composed of ordinary matter, governed by the laws of physics and chemistry; and teleological thinking has been long ago abandoned in physics and chemistry. The world of physics is a world without goals or aims. Therefore, it is not obvious how biological goals arise. In virtue of what do biological systems acquire goals? One answer is that such goals are attributed by external observers. According to this view, although biological organisms are, in essence, complex physical systems, it is illuminating to describe their behavior in terms of goals – *as if* it were subject to goals (Nagel 1961; Schaffner 1993). Ultimately, these goals would depend on the ascriptions of some observer.

However, if one has any hope of accounting for intentionality by appealing to biological goals, those goals cannot be ultimately derived from external attributions. For,

⁹⁷ Of course, the task of showing how all this derivation is even plausible is titanic and full of difficulties. One particular problem is the possibility of evolutionary by-products, that is, traits that are side effects of adaptations – but which lack a biological function themselves. For instance, some authors think that music – as most aesthetic activities – has emerged as an evolutionary by-product, and does not have any biological function (see for instance Pinker, 1997). If this were so, there would be a lot of intentional behaviours and states related to musical activities which would have no biological function whatsoever (nor direct nor derived). In consequence, according to Millikan, there would not be definite goals or contents to be found in these activities. I do not want to defend that music is a by-product, but it is plausible that some human behaviours have arisen as evolutionary by-products – and that, nevertheless, we tend to describe such behaviours as goal-oriented.

if this were so, the intentionality of the external observer would remain unexplained. What is needed is an account according to which biological systems have *observer-independent goals* – that is, goals which do not depend on external attributions (Burge 2009).

There have been several attempts at defending the existence of observer-independent biological goals, while acknowledging that biological organisms are governed by the laws of physics. In this chapter, I will focus on two of these proposals: *evolutionary* approaches (Millikan 1984, 2005; Neander 1991), which argue that the relevant evaluative standards are fixed by the selective history of organisms, and *organizational* views (Christensen and Bickhard 2002; Weber and Varela 2002; Di Paolo 2005; Barandarian and Egbert 2014), which hold that observer-independent evaluative standards arise from the process of self-production and self-maintenance of organisms.

I think that the claims at the heart of evolutionary and organizational accounts of biological goals have tended to be taken for granted without enough scrutiny: more specifically, no convincing reasons have been offered in favour of the view that observer-independent goals emerge either as a result of natural selection or as a result of the self-maintenance of organisms (Schaffner 1993; Boorse 2002). In order to alleviate this shortcoming, I will try to suggest a possible way of justifying such claims, or at least providing some motivation for them. However, after examining this suggestion, I will offer some reasons to remain sceptical that these approaches manage to characterize a satisfactory notion of (observer-independent) biological goals – at least, a notion that can be used in an account of intentionality.

6.3 Proliferation and self-maintenance as evaluative standards

6.3.1 Evolutionary and organizational accounts of biological goals

The most popular accounts of biological goals regard Darwinian evolution as the source of biological normativity. These evolutionary accounts argue that natural selection

introduces an evaluative standard (Millikan 1984, 1989; Neander, 1991).⁹⁸ The basic goal of living systems, on these views, would be to pass the filter of natural selection, that is, to survive and reproduce – in short, to increase biological fitness. Etiological versions of evolutionary approaches (Millikan 1984; Neander 1991) focus on the evolutionary history of organisms: the particular goals of specific biological mechanisms would be to do whatever *past* tokens of the mechanism did that lead to the proliferation of further tokens.⁹⁹

The second perspective on biological goals that I will discuss here may be referred to as *organizational* approach (see, among others, Varela 1997; Weber and Varela 2002; Di Paolo 2005; Christensen and Bickhard 2002; Mossio, Saborido and Moreno 2009; Christensen 2012; Mossio and Bich forthcoming). This approach comprises a broad array of perspectives, which diverge in several respects. However, there are some nuclear tenets shared by most authors working in this framework – in particular, regarding the emergence of biological goals.

Organizational approaches see living organisms as autonomous, self-produced entities that keep their identity amidst a continuous exchange of matter and energy with the environment. An organized system can be said to be *self-maintained* insofar as its own dynamics contribute to sustaining the conditions required for the maintenance of that organized system – if the system does not work in certain ways, those conditions will not obtain, and the system will disappear (Mossio, Saborido and Moreno 2009). For instance, if a protozoon does not ingest nutrients from the environment, it will not be able to regenerate its structure and will eventually disintegrate. In contrast, the continuous existence of a rock does not depend on anything done by the rock itself – the

⁹⁸ Papineau (1987, 1993) also explains biological goals in terms of natural selection, although he takes this to show that such goals are not actually normative standards.

⁹⁹ Propensivist views (Canfield 1964; Ruse 1971; Bigelow & Pargetter 1987), in contrast, focus on the fitness of current organisms: the biological goal of organisms would be to increase their *current* contribution to the proliferation of future tokens of the same type of individuals. In both cases, biological goals are specified in terms of success in a selective regime. For the sake of simplicity, I will focus in etiological versions of evolutionary approaches, although the conclusions I will reach can also be applied to propensivist alternatives.

rock just persists, given suitable environmental conditions (Christensen and Bickhard 2002).

According to the organizational approach, self-maintenance establishes a norm: to “keep the unity going and distinct” (Di Paolo, 2005: 434), that is, to remain within the conditions of viability of the system (Barandiaran and Egbert 2014). The basic goal of living organisms, thus, is the self-preservation of their organization, which is constantly challenged by the disintegrating tendencies of the environment.¹⁰⁰

6.3.2 A challenge for evolutionary and organizational accounts

In this section I will present a challenge that both evolutionary and organizational accounts have to meet. According to these accounts, meeting certain conditions (i.e. proliferating and self-maintaining, respectively) amounts to achieving observer-independent biological goals. But it is not obvious why this is so. Defenders of these accounts must be able to offer reasons for treating either proliferation or self-maintenance as constituting observer-independent evaluative standards. I will discuss this challenge at the same time in relation to evolutionary and organizational accounts, since it applies equally to both approaches – and the possible replies I will consider are analogous in the two cases.

On the face of it, it may seem somewhat arbitrary to posit either proliferation or self-maintenance as observer-independent goals of biological systems. After all, it seems that one could as well say that the goal of living organisms is to disappear and merge with inert matter. Why should the goal of proliferation – or self-maintenance – be privileged over alternative goals like disintegration? Moreover, why should proliferation or self-maintenance be regarded as the achievement of a goal, and not just as the result of certain complex causal processes? Are self-maintenance and proliferation observer-independent biological goals, or rather goals attributed by us as observers, when interpreting non-teleological processes as if they were teleological?

¹⁰⁰ 'To be is its [the organism's] intrinsic goal' (Jonas 1968: 243)

It is true that both proliferation and self-maintenance set observer-independent criteria of demarcation – that is, conditions that are met or not met independently of the attitudes of external observers. Whether some trait proliferates, or whether some organism manages to preserve its organization, is a factual matter that does not depend on the perspective of external observers. However, this does not mean that such criteria have any observer-independent evaluative – or otherwise normative – purport. There are many observer-independent demarcation criteria which lack relevant evaluative implications (at least if such implications are not attributed by some suitable external observer). Think for instance of the property of being a star weighing more than 10^{33} kg. Although having this property is certainly an observer-independent matter, it does not seem to have any evaluative significance – unless relevant external observers start evaluating stars with respect to the possession of such a property (say, they may think that stars heavier than 10^{33} kg are not good for certain scientific research). Otherwise, evaluative standards would proliferate all around: weighing less than 10^{11} kg, more than 1000kg, being composed of carbon – and so on – would all count as evaluative demarcations. I take it as obvious, thus, that not all observer-independent criteria of demarcation set observer-independent evaluative standards (see Hattiangadi 2009; Whiting 2009).

Is there any reason to characterize the demarcation criteria established by natural selection and self-maintenance as evaluative standards? Perhaps it can be argued that regarding proliferation, or self-maintenance, as establishing a goal is extremely useful for biological research – perhaps thinking in terms of such goals is illuminating for our understanding biological organisms and their evolution (Christensen and Bickhard 2002; Christensen 2012; Weber and Varela 2002). This may be so, but it would only mean that it is convenient for certain external observers (i.e. biologists) to attribute such goals to biological systems, not that these systems are subject to those goals in the absence of external attributions.

A stronger version of the argument in the previous paragraph is to claim that treating natural selection, or alternatively self-maintenance, as introducing evaluative standards is not only useful, but *essential* for our understanding of biological organisms and their dynamics (for a defence of this claim in relation to self-maintenance, see Christensen and Bickhard 2002; Weber and Varela 2002; Christensen 2012). Perhaps the

introduction of such standards is an indispensable ingredient of our best biological explanations: according to this argument, if natural selection or self-maintenance were not seen as giving rise to evaluative standards, our explanations of biological phenomena would be significantly impoverished.

It is sometimes argued that if positing some entity is indispensable for our best (scientific) explanations of some phenomena, then such an entity should count as existing (famously, Quine (1980, 1981) and Putnam (1979, 2012) developed arguments along these lines in favour of the existence of mathematical entities). If this is right, the indispensability of evaluative standards associated with natural selection or self-maintenance for biological explanations would warrant belief in the existence of such evaluative standards. But, first, the claim that belief in the existence of an entity is warranted by its explanatory indispensability is controversial (Maddy 1992, 1995, 1997; Sober 1993); and, second, no convincing reasons have been provided for thinking that natural selection and self-maintenance cannot be both suitably explained without appealing to observer-independent evaluative standards.¹⁰¹ On the contrary, it seems plausible to think that the processes involved in natural selection and self-maintenance can be accounted for in terms of causal interactions, without having to posit normative evaluative standards. For instance, natural selection can arguably be explained in terms of the comparative probabilities of proliferation of different biological traits under certain environmental pressures: one does not need to use genuinely evaluative vocabulary in such explanations. At least, it is far from obvious that this sort of non-teleological explanation will be inadequate. Likewise, it seems that one can just say that certain traits or behaviors contribute (or do not contribute) to the self-maintenance of the system – i.e. make the self-maintenance of the system more (or less) probable –, without appealing to evaluative notions.

A further possible argument for regarding proliferation or self-maintenance as evaluative standards is that the very existence of an organism depends on its self-

¹⁰¹ So, rather than being indispensable for science, evaluative notions would be at best pragmatically useful (this reply would be analogous to Hartry Field's objection (1980) that mathematics are not indispensable for science, but merely very useful for it).

maintenance (Christensen and Bickhard 2002; Mossio, Saborido and Moreno 2009; Mossio and Bich 2014) and, analogously, the continuity of a species depends on its proliferation. This argument is exemplified in the following quote:

Because of the organizational closure, the activity of the system has an intrinsic relevance for the system itself, to the extent that its very existence *depends on* the effects of its own activity. Such intrinsic relevance, we hold, generates a naturalized criterion to determine what the system is 'supposed' to do. In fact, the whole system (and its constitutive processes) 'must' behave in a specific way, *otherwise* it would cease to exist. Accordingly, the activity of the system becomes its own norm or, more precisely, the conditions of existence of its constitutive processes and organization are the norms of its own activity. (Mossio, Saborido and Moreno 2009: 825).

A similar idea is expressed by Barandarian and Egbert:

This recursive nature of living organization allows one to derive the norms, not directly from isolated events (as the problem was originally posed by Hume), but from the set of organized and interdependent processes that constitute an organism (i.e., its conditions of possibility). (Barabdarian and Egbert 2014: 10)

However, this will not help, unless it is shown that the preservation of the existence of a system, or the perpetuation of species— rather than, say, the disappearance of living organisms – are *observer-independent* goals, something to aim at. Conditions of existence do not automatically count as setting evaluative standards.

The challenge to evolutionary and organizational specifications of goals that I have been presenting here is not always properly discussed in evolutionary and organizational approaches, although it is mandatory for these approaches to find a suitable answer (Boorse 2002; Scahffner 1993). In the following sections I will suggest a possible way of replying. I will start by presenting this answer in the case of evolutionary accounts, and later I will examine an analogous proposal for the organizational framework.

6.4 Selection and evaluation

I would say that the reason why it seems intuitive to think of Darwinian evolution as setting an observer-independent goal for biological systems is because we are inclined to regard processes of selection as revealing evaluative standards – and evolution is usually understood in terms of natural selection. Selective discrimination preserves some items, performances or traits and discards others: the persistence of the items or performances selected is promoted, while the persistence of those discarded is inhibited. This sort of discriminative reinforcement is typically taken to reveal some underlying evaluative standard. For example, if we see a farmer putting red apples in a basket and throwing brown apples away, we will – prima facie – assume that the farmer is assessing and selecting the apples in accordance with some evaluative standard (say, the farmer may be keeping apples that are good for eating). In general, normative standards are taken to be revealed by the presence of sanctions that reinforce positively certain items/performances and negatively others.¹⁰²

Natural selection may be thought of as a selective process of this kind. Biological systems that accord with the standard set by natural selection will tend to be more frequent in the future, whereas the perpetuation of systems which fail to do so will be inhibited. This is exactly the way in which evaluative selective sanctions are expected to work. Since evaluative sanctions are a typical way in which evaluative attitudes are manifested, it seems that it can be said that, in a sense, the increase of fitness is a biological normative standard in relation to which living systems are evaluated – a standard setting a goal for biological systems.

It may be protested that this notion of goal is, at best, metaphorical, since normative evaluative sanctions, in the paradigmatic cases, are imposed by *agents*, and

¹⁰² I am not saying that normative attitudes are *always* revealed through sanctions, but only that they can be revealed in this way. I am using the term ‘sanction’ in a broad way. The general idea is that behaviors that tend to preserve or promote some state or performance can be seen as revealing positive evaluative attitudes towards such state or performance (while behaviors tending to suppress or inhibit a state or performance would reveal negative evaluative attitudes).

not by impersonal natural forces (is a winter storm evaluating negatively the behavior of the animals freezing in it?). But I will grant – at least for the sake for argument – that it is not gratuitous to talk, in the sense discussed above, of biological goals, given the many 'family resemblances' with the paradigmatic cases – keeping in mind, of course, the peculiarities of this use of such concepts.

A further issue is whether this sort of biological goal may offer grounds for developing a suitable notion of intentionality. I see a difficulty. When the behavior of an organism is interpreted in intentional terms, goals are typically attributed to such an organism *as an individual*. However, the sort of biological goals arising from natural selection seem to be attached to *reproductive chains or lineages linking different token individuals*, rather than to the individuals themselves. Natural selection, as a sanctioning force, often acts *across individuals belonging to different generations*.¹⁰³ The sanctions issued by natural selection, at least in many cases, will not be addressed to the particular individual displaying some behavior, but to its potential descendants – these sanctions will reinforce or inhibit the proliferation of behaviors or traits in the future, typically in *future individuals* which are not the same individual whose performance is being currently evaluated.¹⁰⁴

By contrast, when the behavior of an individual is evaluated with respect to a goal ascribed to that individual *as such* – and not as part of some larger entity –, it is the future performances of *that same individual* which are reinforced or inhibited by the evaluative sanctions. Normative evaluations and sanction should be addressed to the entity which is regarded as subject to the relevant goal or evaluative standard – and not to some

¹⁰³ Mossio and Bich (in press) also note that the goals specified by evolutionary accounts should be ascribed to lineages, although they do not develop the point in detail.

¹⁰⁴ It could be argued that token biological systems or traits may be evaluated as tokens in virtue of belonging to a certain type. So, murder as a type-act is wrong, which makes each token-act of murder wrong. Note however that, still, each token-act of murder is taken to be wrong as a token-act, and that therefore agents will be disposed to address sanctions to the specific individual responsible for a particular token-act of murder. In the case of natural selection, in contrast, it is common that the token-trait that fails to achieve a biological goal is not sanctioned: the only sanctions affect the potential offspring of the organism that token trait belongs to.

different entity.¹⁰⁵ For example, if you win an individual race it is you who should receive a medal, since it is your individual performance what is being evaluated. If the medal were given to some other individual, there would not seem to be reasons to say that the medal is expressing an evaluation of your performance as an individual (say, perhaps you are representing your university in the race, so the medal is given to a representative of the university).

All this suggest that, unless further reasons are provided, the biological goals derived from natural selection should be ascribed to *chains* of biological systems, rather than to individual systems. This may be all right for the purposes of biologists, but as I said above, it is problematic if one wants to ground intentionality on such biological goals. The reason for this is that the sort of goals relevant for intentionality is typically ascribed to *particular agents*. For instance, when some animal is trying to find food, we will say that it is that *individual* animal – and not the reproductive chains it belongs to – who succeeds or fails in finding food.¹⁰⁶

This is not just a mere terminological quibble. The assignment of biological goals made by evolutionary approaches clashes with our intuitive intentional interpretations of the goal-directed behavior of organisms – precisely because the goals identified by these approaches are not goals of individuals. Take the example of a homosexual animal – say, a rabbit – and assume that homosexual behavior has no adaptive function for such animals. Perhaps as a result of some mutation (or for some other reason), this individual rabbit shows clear gratification when engaging in those atypical forms of intercourse,

¹⁰⁵ Assume that sanctioning that other individual is not an indirect way of sanctioning me (you could punish someone indirectly by punishing her friends, since this will presumably upset her).

¹⁰⁶ Maybe the advocate of evolutionary approaches could argue that only those goals in which natural selection operates at the level of the individual (rather than at the level of reproductive chains) are relevant for accounting for intentionality. The problem is that this leaves out many behaviors that we want to be able to account in intentional terms – for example, when parents feed their offspring, or when individuals engage in sexual intercourse.

Furthermore, it may be argued that, according to evolutionary approaches, negative consequences for the survival of a token organism (e.g. the organism being killed by a predator) are only seen as constituting a failure in achieving biological goals insofar as they prevent the proliferation of the type instantiated by that token organism. Thus, the goals revealed by such a process of selection would still be attributable to lineages, rather than to individuals.

whereas failure in doing so brings great displeasure. However, assuming that homosexual intercourse lacks a selected function, advocates of an evolutionary account of goals should accept that, no matter what appearances may suggest, engaging in these forms of intercourse is not a real goal of this individual. What is more, the actual goal of this particular individual would be to engage in heterosexual reproductive intercourse – since this is the selected sexual behavior of this kind of animals.

This conclusion seems to go against our intuitive reading of the example. If the rabbit shows clear gratification after homosexual encounters, and seeks them actively, I take it that the intuitive way of interpreting the rabbit's behavior is to say that it has the goal of engaging in such homosexual encounters. This particular rabbit would be showing clear behavioral signs of having the goal of engaging in homosexual intercourse. This sort of behavioral signs ground our intuitive attributions of intentionality to biological organisms, and it seems hard to disregard them completely.¹⁰⁷

The problem, as explained above, is that the selective pressures associated with natural selection do not always affect the individual whose behavior is being considered. The homosexual rabbit of the example could lead a completely gratifying life – full of successful non-reproductive encounters. There would be nothing in the life *of this individual* that resembles an evaluative sanction. The only drawback is that that individual would have no descendants, but this brings no displeasure (assume that these animals only care about intercourse, but not about the production of offspring). Thus, selective pressures in this example would not have repercussions for the behavior of the individual considered.

It seems, therefore, that in this kind of examples evolutionary approaches make wrong predictions (see Boorse 2002 for related objections). Evolutionary accounts characterize goals which can become too removed from the actual behavioral markers of

¹⁰⁷ See Braddon-Mitchel and Jackson (1997) for a related objection. Braddon-Mitchel and Jackson argue that individuals must have been able to attribute intentional content to agents before the theory of evolution was proposed. Something else than evolutionary history, thus, made agents count as possessing contentful states. These authors claim that, if this is so, it is not necessary nowadays to appeal to evolution in order to characterize what counts as having contentful states – possession of intentional content would keep being characterized nowadays by whatever characterized it before Darwin.

goal-directedness manifest by individuals – that is, their expressions of gratification, displeasure or preference. Of course, it could be argued that our pre-theoretical attributions of goals are systematically misleading – so that the homosexual rabbit above would not after all have the goal of engaging in homosexual intercourse. But I think this is a high price to pay. At any rate, it cannot be accepted in a pragmatist framework which tries to account for intentionality by studying what agents *do* – how they interact with the world.

In conclusion, although an evolutionary approach may capture some interesting notion of biological goal, I do not think that such goals can lay the foundations for an account of intentionality. In the next section I will examine whether organizational accounts are more suitable for this task.

6.5 Self-maintenance as the source of teleology

The reasons for privileging the goal of self-maintenance – and not, for instance, self-destruction – parallel those that were provided for proliferation as an intrinsic goal in evolutionary approaches. Trivially, systems that preserve their identity through interactions with the environment will keep existing, while systems that fail to maintain themselves will most likely disintegrate and cease to exist. Those processes and performances that contribute to self-maintenance are reinforced, whereas those that do not tend to be inhibited. Therefore, interactions with the environment can be seen as *sanctioning* the behavior of living organisms with respect to the normative standard of self-maintenance. As in the case of evolutionary proposals, these sanctions may be regarded as implicitly expressing normative evaluations – albeit, again, in an extended, almost metaphorical way.

Note that this organizational proposal seems to be free from the objection posed above to evolutionary approaches – that is, the objection that biological goals associated with natural selection should be ascribed to reproductive chains, rather than to particular individuals. According to the organizational view – as I have been describing it –, when a particular system succeeds in accomplishing the goal of self-maintenance, it is the future existence *of that same system* which is promoted (not the existence of descendants or

other individuals). Actually, the key point of organizational approaches is that organisms aim at preserving *their identity* by self-preserving their organization and boundaries. Therefore, the biological goals arising in this organizational framework are goals of individuals – as is demanded by intentional explanations.

However, a related problem comes to view: organizational approaches seem to be unable to account for biological goals related to reproduction. It is common in nature that organisms do things with the aim of securing reproduction – or, more generally, of securing genetic spread –, even at the cost of their own self-maintenance. For instance, animals often devote a lot of resources to feeding and protecting their offspring. Organizational proposals, at least without further sophistications, cannot explain these behaviors and the biological goals underlying them, since such behaviors do not contribute to the self-maintenance of the organism (Mossio, Saborido and Moreno 2009; Delancey 2006).

A possible solution is to extend the system to which self-maintenance applies: maybe some biological goals are attached to the self-maintenance of reproductive families or lineages (Christensen and Bickhard 2002; Saborido, Mossion and Moreno 2011). In these solutions, however, the focus on individual organisms and their dynamics is lost, and, therefore, the objection to evolutionary accounts introduced above would come back into play (these goals should be attributed to such larger entities, i.e., lineages, and not to individuals). As a result, organizational views seem unable to provide a suitable characterization of many of the goal-oriented behaviors typical of intentional agents.¹⁰⁸

Before finishing this section, let me present a line of development of the organizational framework which is related to the view I will propose in chapter 8. Some authors working from an organizational perspective, in particular in enactivist approaches (Di Paolo 2005), argue that biological systems only start developing something that may be seen as a minimal form of intentionality when they become responsive to their own goals – in particular, to the goal of self-maintenance. According to these authors, such

¹⁰⁸ Although I will not discuss the issue here, I also think that the organizational approach faces serious difficulties in relation to goal-directed behaviors that, nonetheless, seem to go against the self-maintenance of the agent – for instance, suicidal or unhealthy behaviors (smoking, eating too much sugar), or selfless sacrifices for the sake of others. See Di Paolo (2005) for a tentative response.

responsiveness emerges when organisms become able to *evaluate* (fallibly) their own state in relation to the goal of self-maintenance, and to regulate it in suitable ways (Di Paolo 2005; Barandiaran and Egbert 2014).

The idea is the following. Biological systems have a range of viability – that is, a range of states in which the organization of the system can be preserved. Systems sensitive to their own self-maintenance would monitor their state so as to keep it within such a range, and would try to counterbalance perturbations driving the system out of the viability zone (Di Paolo 2005; Barandiaran and Egbert 2014). This kind of self-regulation is called by Di Paolo 'adaptivity' (2005). Adaptivity, according to Di Paolo, makes the organism responsive to its own goals, and would give rise to the minimal intentionality that he wishes to attribute to living organisms. So, for instance, a protozoon that regulates its movement so as to go to areas with high concentrations of nutrients can be said to be *looking for* nutrients– the behavior of the protozoon is governed by such a goal.

A further possibility is that the goals of an organism are determined by its self-regulating behavior – by the range of states the system tries to remain in. However, both Di Paolo and Barandarian and Egbert seem to think that the biological goals of the organism are set by its conditions of self-maintenance. Adaptivity (i.e., self-regulation) would make the system's behavior sensitive to its own goals – allowing for the sort of responsiveness demanded by intentionality –, but would not determine such goals. On Di Paolo's view (2005), observer-independent biological goals are ultimately established by the process of self-production and self-maintenance of the system. Therefore, it would be in principle possible that the whole self-regulatory activity of an organism functioned incorrectly with respect to the goal set by self-maintenance (although of course, it is not likely that the organism would manage to maintain itself in such conditions).

Many of the problems faced by organizational (and also evolutionary) accounts of biological goals would be overcome, I think, if evaluative standards were taken to be introduced by the regulatory behavior of agents, rather than by self-maintenance (or by natural selection). In chapter 8 I will explore this view.

6.6 Conclusions

In this chapter I have examined two accounts of biological goals – evolutionary and organizational accounts – and discussed whether they offer a suitable characterization of the conditions of success of actions. My main objection to both approaches has been that their assignation of goals is too detached from the behavior of the agents subject to such goals – something that clashes with our intuitive attributions of intentionality.

In particular, the goals specified by evolutionary accounts seem to be attached to reproductive chains, rather than individual agents. However, when we interpret the behavior of an agent in intentional terms, we typically attribute goals to that individual agent (not to cross-generational chains). Organizational approaches, on the other hand, seem able to specify goals of individuals, but only at the cost of disregarding goals that have to do with the reproductive aims of agents, rather than with their own self-maintenance. Again, this last kind of goals plays a central role in our characterization of individuals as rational agents.

In chapter 8 I will put forward an account of goals which overcomes this problems. On the view I will propose, evaluative standards arise by virtue of the regulatory behavior of the agents that are subject to such standards. As a result, the behavior of agents and the goals attributed to them cannot be radically decoupled. Furthermore, in this view evaluations are seen as issued by agents, rather than impersonal environmental pressures.

Before presenting my account of goals, I will discuss in chapter 7 how the problem of ignored obstacles (introduced in chapter 5 in relation to success semantics) can be dealt with within the framework I am proposing.

Chapter 7: The Defeasibility of Practical Reasoning

7.1 Introduction

In chapter 4 I described the role of action and perception in an inferentialist model of reasoning and communication. In particular, I showed how individuals engage with the world through action-perception feedback loops. In short, the idea is that when an agent accepts some beliefs, she takes them as a guide for action: she becomes disposed to act in certain ways so as to achieve her goals. Entitlement to such beliefs will count as providing entitlement to act in those ways, given the agent's goals. When the performance of an action (as a sufficient means to the agent's goal) is properly entitled, I have argued, it cannot fail in achieving its goal: practical failure reveals that the action was not properly entitled. In many cases, the outcomes of the agent's actions depend on how things are in the world. Agents will evaluate these outcomes (typically, through perception) and will revise and update their beliefs accordingly. In this way, the world can be seen as shaping and constraining the cognitive views of agents.

So, discursive practices become sensitive to the way the world is because the results of interactions with the environment are determined, at least in part, by how things are in the world, and agents modify and update their views in accordance with their evaluations of these results. In virtue of agents' practical engagement with their environment, discursive practices can be said to answer to how things stand in the world – which allows for the emergence of a representational dimension in such practices.

This model sketches a natural way of constructing a pragmatist account of representation. Peirce's thesis that the content of beliefs is given by their practical consequences finds an obvious development in the idea that what these beliefs represent is states of the world which make such practical endeavours successful: states of the world in which those beliefs are good practical guides for achieving the agent's goals. In chapter 5, I discussed a straightforward rendering of this sort of pragmatist approach, success semantics. Now, I will examine how this pragmatist project may be fleshed out in the inferentialist framework I outlined in the first chapters. In particular, I will discuss how this pragmatist approach may deal with the difficulties faced by success semantics. I will focus on what I take to be the two main problems for this kind of views. First, in this

chapter, I will consider the problem of ignored obstacles for the success of actions based only on true beliefs. Second, in the following chapter, I will address the issue of the assessment of the success of our actions – and of how the world may influence such assessment.

My main contention in this chapter will be that the possible presence of an indefinitely large number of ignored obstacles makes practical reasoning defeasible – in the sense that a piece of practical reasoning recommending some action as a good means for an end may become bad due to the presence of an ignored obstacle for the achievement of the end. I will claim that such defeasibility introduces a form of particularism about practical reasons: whether some consideration offers good reasons for performing some action as a means to an end does not only depend on the truth of such a consideration, but also on whether ignored obstacles obtain. As a result, it is not possible to identify directly the truth conditions of particular beliefs with the success conditions of some set of actions, as success semantics tries to do. This does not mean, however, that the success of our actions does not depend on the truth of our beliefs, but only that such dependence is not piecemeal. Thus, the updating of our beliefs in the face of practical failures answers to the world, but in a holistic way. Practical failure reveals that there is something wrong in our views about the world; however, whether a specific belief should be revised in light of the failure of some specific action will depend on what other beliefs are held by the agent and on the practical outcomes of related actions. This does not make the revision of our views indeterminate – when deciding how to update our beliefs in a particular situation, in general some specific updates will be clearly more reasonable than the alternatives (given the circumstances surrounding that particular situation, for instance the outcomes of previous actions or what obstacles are likely to obtain in those circumstances).

My aim in this chapter, therefore, will not be to overcome the holism introduced by the possibility of ignored obstacles, but rather to show how we can live with such holism – more specifically, I will try to show that despite this holism, it makes sense to say that our practical engagement with the world makes our beliefs answer to how things are, even if not in a piecemeal way.

In the course of this investigation, I will explore the nature of defeasible reasoning (i.e. reasoning subject to exceptions, which can be defeated by the addition of new

information). In particular, I will conclude that defeasible reasoning is occasion-sensitive, in the sense that whether some piece of reasoning is (defeasibly) good is an occasion-sensitive matter. I will argue that which possible defeaters – which possible exceptions – have to be explicitly ruled out by the premises of an inference and which possible defeaters may remain unconsidered (without making the inference enthymematic) depends on different features of the occasion of inference – including what conditions are typical in the context, what information is available or how high stakes are. I will argue that possible defeaters need to be explicitly discarded by the premises of non-enthymematic pieces of reasoning only when they are relevant enough (that is, only when there are reasons enough for the agent to suspect that such defeaters might obtain); in turn, whether a given defeating possibility is relevant enough on a given occasion of use will be determined by different features of the circumstances surrounding that occasion of inference. Interestingly, this introduces an occasion-sensitive dimension in the sort of inferentialist semantic theories I have been considering, since according to such theories the content of claims is determined in part by the defeasible inferences they are involved in.

The structure of the chapter is as follows. In section 7.2 I examine how the possibility of ignored obstacles is incorporated into an inferentialist framework. In section 7.3 I relate the problem of ignored obstacles to the defeasibility of practical reasoning. In section 7.4 I put forward a view of defeasible reasoning which makes it possible to develop an inferentialist-pragmatist account of representation. In section 7.5 I discuss the problem of ignored obstacles from the perspective of this view of defeasibility. After that, in section 7.6 I argue that defeasibility introduces an occasion-sensitive dimension in inferentialist approaches. In section 7.7 I recapitulate by presenting the picture of representational practices resulting from the discussion in the previous sections.

7.2 Ignored obstacles

As I explained in chapter 5, the success of an action does not seem to be guaranteed by the truth of the set of beliefs that license it. There may be ignored facts that thwart the success of the action, even when the agent has based her action only on true beliefs

(Brandom 1994b). Accepting Brandom's contention that, in most cases, the number of possible obstacles is indefinitely large, it seems implausible to attribute to agents beliefs about the absence of all these obstacles. This is a problem for versions of success semantics that hold that truth is sufficient for practical success (Whyte 1990; Dokic and Engel 2002). Furthermore, even if it were granted that agents have (implicit) beliefs ruling out the presence of all obstacles – so that the truth of the agent's beliefs guarantees practical success –, it will be problematic to detach the contribution of all these (indefinitely many) anti-obstacle beliefs to the success-conditions of actions. Thus, the truth conditions of individual beliefs do not seem to be derivable from the success-conditions of actions (since these success-conditions include an indefinitely large number of anti-obstacle conditions, which cannot be fully detached).

At first sight, there is no special difficulty in integrating the possibility of ignored obstacles in the inferentialist framework I have presented. Remember that, in the account of practical reasoning I outlined in chapter 4, when an agent endorses a piece of practical reasoning, she takes the premises to offer good reasons for performing a certain action – in view of the goal that underpins her endorsement of the piece of reasoning. I focused on means-end instances of practical reasoning, in which some action is recommended as a good means for achieving the goal of the agent. So, for an agent that wants to cook pasta (she is hungry), the fact that the water in the pot is boiling provides good reasons for performing the action of putting the pasta in the pot. This is so because putting the pasta in the pot will lead to the pasta being cooked, given that the water in the pot is boiling. Thus, the inferential link represented by (1) is good:

(1)

It is desirable for me to cook the pasta

The water in the pot is boiling

I ought to put the pasta in the pot

The goodness of a given means-end piece of practical reasoning depends on the goodness of a related theoretical inference according to which the recommended action is a good means for achieving the agent's goal. I will call this theoretical inference, instrumental inference – since it presents some action as an instrument for producing certain results. The destinies of the instrumental theoretical inference and the corresponding piece of practical reasoning are linked: if the theoretical inference is not good, the associated piece of practical reasoning is not, either. So, in order for (1) above to be good, the following theoretical inference must be good.

(2)

The water in the pot is boiling

I put the pasta in the pot

The pasta will get cooked

If the agent thinks that (2) fails to be good – for whatever reason –, she will also be committed to not accepting the associated piece of practical reasoning (1). The dependence of the goodness of means-end patterns of practical reasoning on the goodness of instrumental theoretical inferences can be made explicit – by agents with enough expressive resources – by including among the premises of the pattern of practical reasoning a conditional codifying the instrumental inference (as I already pointed out in section 4.4). So, (2) is codified by the conditional premise added in (3):

(3)

It is desirable for me to cook the pasta

The water in the pot is boiling

If the water in the pot is boiling and I put the pasta in the pot, the pasta will get cooked

I ought to put the pasta in the pot

A crucial point is that instrumental theoretical inferences are typically defeasible (i.e. non-monotonic). That is, the addition of new premises can make the inference stop being good. For instance, (2) becomes bad if the pressure is very low (boiling temperature will also be low), or if there is some substance in the water that prevents pasta from getting cooked. This is the kind of ignored obstacles that was a problem for success semantics. In this inferentialist framework, the effect of such obstacles is to defeat the instrumental theoretical inference connecting means and ends – so that the conditional that codifies it becomes false. Accordingly, the corresponding piece of practical reasoning gets defeated as well.

The introduction of one of these obstacles will result in the failure of the action which was recommended in the absence of such obstacles. Thus, the agent will not actually have good reasons for performing that action (as a means to her end). Of course, if she is not aware of the presence of the obstacle she may be epistemically blameless (maybe there were no reasons to suspect the presence of the obstacle). But, anyway, she will be doing something inappropriate: she will be undertaking a practical commitment (e.g., to putting the pasta in the pot) to which she is not properly entitled, given her goals and the state of the world.

Therefore, the failure of the agent's action can be traced back to something inappropriate in her cognitive views. Even if all her beliefs are correctly held, she may still be endorsing – perhaps implicitly – a bad practical inference: she may be taking those beliefs to license some action that is not actually properly licensed by them – due to the presence of some defeating obstacle. One may wonder whether endorsing a bad inference amounts to having some false belief. Perhaps this is so, when the endorsement of the inference is made explicit by accepting the conditional that codifies it: the agent would have a false conditional belief (all this, assuming that there are conditional beliefs). I do not think much hinges on this, however. At any rate, it can be said that there is something incorrect in the agent's cognitive views – be it a false belief or an

inappropriately endorsed inferential commitment –, and this accounts for the failure of the action.

So, the possibility of an indefinitely large number of possible obstacles is integrated in the inferentialist model by making theoretical inferences connecting means and ends – and the associated pieces of practical reasoning – defeasible. In this way, the inferentialist model I am presenting can handle the possibility of ignored obstacles, something that success semantics was unable to do (at least, simple versions of success semantics). However, this greater flexibility is not without problems. It may be thought that the connection between practical success and the accuracy of the agent's beliefs becomes too loose, so that it cannot determine the representational purport of such beliefs. Practical failure may be due either to the falsity of the agent's beliefs or to ignored obstacles; thus, the conditions of success of the actions elicited by some beliefs do not provide enough information to determine at the same time the truth conditions of those beliefs and the conditions under which such actions are properly licensed by them – in other words, the conditions of success of actions do not allow us to fix both the truth conditions of the agent's non-conditional beliefs and the truth conditions of the conditional beliefs associated with the relevant piece of practical reasoning. In the next section I will address this indeterminacy charge.

7.3 Ignored obstacles and the problem of indeterminacy

It should be noted that the sort of indeterminacy I have referred to above would not arise if practical reasoning were not defeasible – that is, if the associated instrumental theoretical inference were monotonic, so that no ignored fact could defeat it. True, in such a case the success of the agent's action would still depend on whether the premises of her piece of reasoning properly licensed the action performed – that is, on the validity of her reasoning, in addition to its soundness.¹⁰⁹ Nevertheless, the validity of the piece of practical reasoning would not depend on the circumstances in which the action is

¹⁰⁹ Remember that in the present inferentialist framework, an inference may be (materially) valid without being deductively valid. An inference will be valid if and only if entitlement to the premises provides entitlement to the conclusions (see note 63, in section 4.4).

performed: it would be either invariantly valid or invariantly invalid. Of course, such circumstances would determine whether the reasoning was sound (whether the agent's beliefs were correctly held), but not whether entitlement to those beliefs provides entitlement to perform the action. So, the monotonic inferential link between "Madrid is to the South of London" to "London is to the North of Madrid" is valid no matter what conditions the agent finds herself in (of course, if Madrid were actually to the North of London, the inference would be unsound).¹¹⁰ If practical reasoning were monotonic, the success of the actions recommended by pieces of practical reasoning would only depend on whether such pieces of reasoning were sound in those circumstances – whether the premises were accurate.

The defeasibility of practical reasoning changes all this. Given some fixed goals of the agent, a certain piece of practical reasoning would be good in some circumstances and bad (i.e. defeated) in others – in circumstances where defeaters are present. In the same way, the associated instrumental inference would be valid in some circumstances and invalid in others. So now the conditions of success of actions would carry information about the conditions in which the beliefs held by the agent were accurate *and* the piece of reasoning employed was good. And, in general, the correctness conditions of a conjunction do not provide enough information to factor out the correctness conditions of each conjunct (unless it is known that one of them is always correct).¹¹¹

One way of removing the defeasibility of the rational connection between the premises and the action recommended is to add a conditional premise saying that, if the rest of premises hold, then the recommended action will be successful – as in (3) above. This move is analogous to Whyte's (1990) introduction of a no impediments belief (a belief to the effect that there are no obstacles, so that the agent's actions will be successful, provided that her other beliefs are true).

¹¹⁰ What determines whether a given piece of reasoning is valid? In the framework I am favoring here, the claims involved in such piece of reasoning. In turn, what claims are expressed by certain discursive moves is determined by the norms of the practice.

¹¹¹ This would be the case with a monotonic valid piece of practical reasoning: the rational link between premises and practical conclusion would always be correct. In contrast, the premises would only be accurate in some conditions.

With the addition of the conditional premise, the instrumental theoretical inference becomes then formally good: its validity does not depend on the content of the non-logical concepts involved. In the same sense of ‘formal’, the associated practical inference (3) will be formally good (regarding ‘desirable’ and ‘ought’ as logical concepts, in an extended sense).

(3)

It is desirable for me to cook the pasta

The water in the pot is boiling

If the water in the pot is boiling and I put the pasta in the pot, the pasta will get cooked

I ought to put the pasta in the pot

In this formulation, the conditions of success of the action recommended would only depend on the conditions of accuracy of the premises; the rational link between premises and conclusions is now (indefeasibly) valid in all circumstances. However, all this is at the cost of introducing a new conditional premise: the conditions of success will now reveal the truth conditions of the conjunction of the new conditional premise and also the other premises. If the action is unsuccessful it may be either because the non-conditional or the conditional premises are inaccurate (e.g., maybe the water was not boiling, or maybe the conditional premises was falsified by the presence of some defeating obstacle). So, the indeterminacy generated by the defeasibility of the rational connection between premises and the recommendation of the action has been transferred to indeterminacy generated by the presence of a new premise.

If one tries to detach the truth conditions of the additional conditional premise – so as to single out the truth conditions of the non-conditional premises –, the general problem of success semantics, discussed in sections 5.4 and 5.5, will reappear. The

possibility of an indefinitely large number of defeating obstacles will prevent such conditional premises from being effectively detached.

Let me rehearse again briefly how this problem arises. The strategy for detaching the conditional beliefs would be to contrast pieces of practical reasoning in which the same non-conditional premises $\{B_i\}$ lead to different actions A_j . The conditional instrumental beliefs ($\{B_i\} \rightarrow \text{success } A_j$) will also be different in each of these pieces of practical reasoning. So, by comparing the conditions of success of all the actions A_j , one could hope to single out the truth conditions of the beliefs $\{B_i\}$ (which constitute the common premises of all the pieces of practical reasoning considered).

However, no matter how many different actions following from $\{B_i\}$ one compares, there always remains the possibility that there is some further common obstacle for all those actions – and, therefore, some anti-defeater condition that has not been factored out. Anti-defeater conditions, thus, mingle with the success conditions of actions in seemingly inextricable ways (for the details, see section 5.5).

In conclusion, I am skeptical that there are good prospects for the project of singling out the truth conditions of individual beliefs from the success-conditions of the actions derived from them – regardless of whether the agent's endorsement of the rational link between the premises and the action recommended is shown explicitly by means of an additional conditional premise. The defeasibility of means-end practical reasoning poses seemingly insurmountable difficulties for this sort of project, both in the form undertaken in success semantics and in the kind of inferentialist framework I am exploring.

However, I still think that there is room for defending the pragmatist idea that the world gets to have a say about the correctness of our views by determining whether the actions that follow from such views are successful. Developing this idea requires to examine with some care the defeasible nature of our grasp on the world. This is what I will do in the next sections.

7.4 Defeasible reasoning

The crucial point discussed in the previous sections is that means-end practical reasoning is underpinned by non-monotonic (i.e. defeasible) inferences. As a result, the correctness of a rational transition between certain premises and the recommendation of an action may be defeated by the presence of some obstacle.

Thus, when the performance of an action fails to fulfill the goals of the agent, it may be either because the beliefs that recommended such an action were not properly held, or because some obstacle defeated the rational connection between these beliefs and the recommendation of the action. In principle, the presence of any specific obstacle may be investigated by exploring the success of further actions derived from the belief that such an obstacle is absent (or present). Suppose, for example, that I suspect that my computer's internet is not working because the lights in my building went out. If this is the case, then the light in the kitchen should not turn on when I press the switch.

However, I am working under the assumption that there are an indefinitely large number of possible obstacles for any practical plan. Thus, it is not possible to discard the presence of every possible obstacle for the success of an action. When an action fails, there is always the possibility – at least in principle – that the cause of the failure is some further unexpected obstacle, rather than the inaccuracy of the agent's beliefs. In other words, the non-monotonic character of practical reasoning cannot be revoked by adding collateral premises discarding the presence of each possible defeater.

On the face of it, one might think that defeasibility prevents agents from being attributed entitlement to accepting the conclusions of their reasoning – since such attributions are never certain, and can always be undermined by the presence of some ignored defeater. This pessimistic outlook, nevertheless, is based on a misleading conception of non-monotonicity – a conception which imposes too exacting demands on defeasible reasoning. In this section, I will present a more plausible view of defeasible reasoning, which makes it possible to develop a pragmatist account of representation.

On the view of defeasible reasoning I will work with, it is acknowledged that attributions of entitlement are provisional and revocable – they are permanently subject to revision. When an agent has good (albeit defeasible) reasons for accepting some

conclusions, she may be attributed entitlement to do so. If it eventually turned out that there is some defeater active, the attribution of entitlement would (or at least ought to) be withdrawn: the agent would be seen as not having had *actually* good reasons for endorsing the conclusions, even if it looked as if she did. But this does not mean that, as long as no defeater is detected, the agent cannot be attributed full entitlement to such conclusions.

In general, reasoning in accordance with some defeasible inferential connection is only possible if agents can be seen as having good reasons for endorsing certain conclusions even if not all possible defeaters for the goodness of the inference have been explicitly ruled out. Otherwise, the notion of defeasible reasoning would not leave the ground: if all possible defeaters needed to be explicitly discarded, then the piece of reasoning would become indefeasible. So, in defeasible reasoning an agent may be entitled to some conclusions without having to give reasons supporting the absence of all possible defeaters. It is enough, I submit, if the agent has *no good reasons* for accepting the presence of a defeater, or for suspecting that a defeater may obtain. Thus, defeasible reasoning requires *absence of reasons* for thinking that a defeater may obtain, rather than positive reasons ruling out that the defeater obtains. In the absence of such reasons, the agent's inference will be seen as good. Of course, the goodness of the inference is still defeasible: if there actually were good reasons to suspect that defeaters obtained (even if the agent did not possess such reasons), her inference would be revealed to be actually bad. But, when there actually are no reasons to suspect that defeaters may obtain, the agent does not need (in order for her inference to be good) to discard their presence explicitly – she does not need to provide positive reasons for believing that no possible defeater is taking place, in order for her inference not to be enthymematic.

So, as long as the agent does not possess reasons to suspect that some defeater might obtain, she may take herself to be entitled to endorsing the conclusions of her piece of reasoning; however, this entitlement will be defeasible: it will be revoked if it turns out that there actually were good reasons for thinking that some defeater might obtain (e.g. if the defeater did actually obtain, or its obtaining was actually likely).

If the difference between having positive reasons for believing *not-p* and absence of reasons for believing *p* is obliterated, there would be no distinction between non-monotonic and monotonic reasoning. This is so because agents with *no reasons to believe*

that there is some defeater active would count as *having reasons to believe* that there are no defeaters. Therefore, these agents could be seen as engaging in monotonic reasoning, with additional (perhaps hidden) premises stating that there are no defeaters. The presence of a defeater would only make the piece of reasoning unsound, by falsifying one of the premises (i.e., there were actually no good reasons for excluding the defeater).

Think, as an example, of the following inference:

(4)

Tweetie is a bird

Tweetie is not a penguin (or a member of any other species of flightless birds)

Tweetie can fly

This inference will be defeated if Tweetie has a broken wing. This does not mean, however, that the agent's reasoning always needs to include an additional premise stating that Tweetie's wings are not broken. It is enough if there is not any good reason for the agent to think that this is may be the case.

The main characteristic of defeasible reasoning is that agents reason in the absence of total information. In standard contexts, agents are entitled to reason under the assumption that there are no defeaters; when there appear reasons for doubting the absence of some defeater, the context ceases to be standard, and agents have to find reasons for discarding the presence of those defeaters. If there are reasons to suspect that a given defeater may be present and, nonetheless, the agent does not rule it out explicitly, her reasoning will be bad – more precisely, it will be enthymematic: she should have given reasons for discarding the presence of that defeater.

I will say that a defeating consideration is *relevant* when there are reasons to suspect that such a defeater may obtain (see McKenna 2014): the stronger the reasons for suspecting that a defeater may obtain are, the more relevant that defeating

consideration is. When a defeating consideration is relevant enough (i.e. there are reasons to suspect that the defeater may obtain), the agent ought to vindicate her entitlement to discarding it before counting as entitled to endorsing the conclusions of her piece of reasoning.

What sorts of facts are reasons for suspecting that a given defeater may obtain? Roughly, facts that increase the likelihood of the presence of the defeater or that make it rationally expectable. It is sometimes claimed that the possibility of a given defeater may become relevant merely by virtue of its being salient – for instance, by being mentioned in the conversation (see Lewis 1996). On the view I will endorse here, however, the fact that some defeating condition is salient does not mean that there are good reasons for suspecting that it may actually obtain. For instance, the fact that a defeating possibility is mentioned in a conversation (perhaps in a careless way) does not need to make its obtaining likely or rationally expectable.¹¹²

Note that defeating conditions can be made relevant by facts ignored by the agent. Suppose, for instance, that an agent is in an island where penguins are typical, but she does not know that this is the case. Still, the possibility that Tweetie (a bird from the island) is a penguin is made relevant by the fact that penguins are actually typical in the island: this is a likely possibility, so unless the agent rules it out, she will not be entitled to concluding that, since Tweetie is a bird, Tweetie can fly. There are reasons to suspect that Tweetie, being a bird from the island, is a penguin – even if the agent does not possess such reasons, in the sense that she is ignorant about the facts that constitute those reasons.

Likewise, the fact that some barn facade is located in fake barn county makes the possibility of its being a fake barn facade relevant (it is a reason to suspect that such a

¹¹² Blome-Tillmann (2009) argues that the possibilities that may remain ignored are those that are excluded by what is pragmatically presupposed by the relevant speakers. According to Blome-Tillmann (2009: 256), some proposition p would be pragmatically presupposed by an agent x in a given context C iff the agent 'is disposed to behave, in her use of language, as if she believed p to be common ground in C ', where the common ground is characterized in Stalnaker's terms (Stalnaker 2002). Merely attending to some defeating possibility does not make it relevant, on this view, unless it is not ruled out by the pragmatic presuppositions of the context.

On the proposal I am putting forward, in contrast, it is not always the case that a defeating possibility may be ignored if it is excluded by the pragmatic presuppositions of the context; it also required that that those anti-defeater presuppositions are *properly* presupposed – there must be no good reason to suspect that the defeating condition excluded may obtain.

possibility may obtain), even if the reasoning agent does not know that fake barns are common in the county. When talking about a facade located in fake barn county, an inference from the premise “There is a barn-like façade in the field” to the conclusion “There is a barn in the field” is enthymematic: for the inference to be (defeasibly) good, it is required that an additional premise rules out the possibility of its being a fake barn facade. This is so regardless of whether the agent making the inference knows that fake barns are common in that county.

Another way of putting this point is saying that it is not enough to be right out of luck: it is not enough that one’s piece of reasoning is not defeated just because one was lucky.¹¹³ In order to count as properly entitled to endorsing the piece of reasoning, one must be able to rule out defeaters whose presence is actually likely – that is, defeating considerations that are actually relevant.

It is plausible to think, moreover, that the relevance of some defeating possibility depends on the information available to agents and on their background knowledge – since such information can make the defeating condition rationally expectable. Imagine you are in an island full of snakes –which, as a matter of fact, are very rarely venomous. Suppose you do not have specific knowledge about the fauna of that particular island, but you know that it is surrounded by other islands where venomous snakes are typical; arguably, you will have good reasons to *suspect* that the snakes in the island may be venomous – since snakes in the surrounding islands typically are. That the snakes in the island are venomous is something actually expectable, given your background knowledge. However, if you are familiar with the specific fauna of the island, it seems there will be no reasons for you to suspect that snakes there are venomous: that possibility is not actually likely or something to expect, given this enriched background knowledge. From the perspective of agents with less background evidence, some possibility can be actually more likely (and thereby more relevant) than it actually is from a perspective with more background evidence.

A further plausible view is that *false or misleading* information can provide *good* reasons to *suspect* that a defeater may obtain. When the information comes from a

¹¹³ In terms of possible worlds, the idea is that a defeating possibility has to be considered if it obtains in nearby worlds. Lewis (1996) captures this requisite in his Rule of Resemblance, according to which worlds close enough to the actual one are always relevant.

usually reliable source, the agent will have good reasons to take it into consideration. For instance, reading in a generally reliable encyclopedia that most snakes in the island are venomous gives me good reasons to suspect that Kaa, a snake from that island, may be venomous – even if the information in the encyclopedia happens to be mistaken and snakes in that island are never venomous. Since the encyclopedia is in general reliable, its (false) information about the island’s snakes makes the possibility that Kaa is venomous relevant. Arguably, the information in the encyclopedia does not give actually good reasons to believe that venomous snakes are typical in the island, or that Kaa is venomous (it is, after all, misleading information); but it does give good reasons to *consider* such a possibility, to regard it as relevant.

Summing up, defeasible pieces of reasoning become bad when an (undefeated) defeating condition actually obtains – regardless of whether the reasoning agent is aware of the presence of the defeater. However, in order for the agent to be entitled to endorsing the conclusions of some piece of defeasible reasoning it is not required that she gives reasons to rule out each possible defeater. A possible defeater only needs to be explicitly ruled out if there are good reasons to suspect that it may obtain (i.e. facts that make its possibility relevant): in that case, a piece of reasoning that does not include a premise discarding it will be enthymematic. When the agent does not possess reasons to suspect that defeaters may obtain (i.e. she knows no facts that make the possibility of defeaters relevant), she may take herself as entitled to endorsing the conclusions of (good) pieces of defeasible reasoning – although such an entitlement is still defeasible and would be revoked if there appeared reasons for suspecting that defeaters may obtain.

7.5 Defeasibility and practical reasoning

The picture of defeasible reasoning I have presented offers a promising way of developing a pragmatist account of representation. An agent can be seen as properly entitled to the performance of an action (as a means for some goal) despite not having explicitly ruled out the presence of possible obstacles for the success of such an action. In circumstances in which there are no positive reasons suggesting that some obstacle might be present, it

will be considered that the agent is entitled – by default – to reason as if such obstacles were absent (see Williams 2001, for a similar view in relation to attributions of knowledge). However, this entitlement can be revised. If there appears some evidence giving reasons to doubt the absence of obstacles (the failure of the action may give rise to this sort of reasons), the agent will have either to provide reasons to discard the presence of those obstacles or to renounce her entitlement to the performance of the action.

The important point is that agents will not need to rule out the presence of every possible obstacle *all at once*. This is the kind of endeavor made impossible by the non-monotonicity of practical reasoning. In general, some possible obstacles will be more relevant than others; there will be stronger reasons to consider the presence of certain obstacles among the many possible ones, given the evidence available – i.e. taking into account considerations about prior occurrences of the obstacle or what conditions are typical in the context. For instance, if you live in a very cold region, when constructing a house it will be relevant to consider the possibility of pipes getting frozen, while this possibility will be far-fetched and can remain unconsidered in warm lands.

When there appears some reason to suspect that the success of the action was thwarted by some unexpected impediment, the agent will only need to investigate *relevant enough* possible obstacles. There will be further possible (at least, possible in principle) obstacles that do not have to be discarded explicitly – unless additional reasons for considering them arise. As I have explained earlier, investigating the presence of specific obstacles does not bring special problems: the agent can just assess the practical consequences of beliefs about the presence or absence of such obstacles. What was problematic was to rule out the *whole* (indefinitely large) list of possible impediments for a given plan.

When an action fails, the agent will have to decide whether the failure was most likely due to the inaccuracy of the beliefs on which the action was based, or to some unexpected obstacle. In many cases, there will be no relevant possible obstacle, so the failure will be blamed – *prima facie* – to some inaccuracy in the agent's beliefs. For instance, if I get burnt when drinking my soup, I will conclude that my belief that the soup is merely warm was not accurate. In principle, the action of drinking without being burnt could have been thwarted by some unexpected obstacle (perhaps the soup has been contaminated with some ingredient that reacts with human saliva producing burns), but

in this context these alternative possibilities are not relevant enough and do not need to be considered – there are no reasons to suspect that such possibilities may have obtained.¹¹⁴

In other cases, the failure of the action may make the possibility of obstacles become relevant. For instance, my confidence in the accuracy of my beliefs (they have led to nothing but success so far) may give reasons to suspect the presence of some unexpected impediment. When this is so, the agent will have to investigate those obstacles whose possibility is more relevant in the context. If all relevant obstacles are discarded (by testing the practical consequences of their presence), the agent may be eventually compelled to revise the beliefs constituting the premises of her reasoning. Of course, this process of revision may be rather complicated – involving examinations of further practical consequences of the agent's premises, as well as practical tests of the presence of obstacles.

Let me present an example. Imagine that I want to speak with a friend on the phone, so I dial (what I take to be) her number. However, some stranger answers. One possibility is that my belief that that is my friend's number is wrong. An alternative explanation of the stranger's answering the phone is that he has borrowed it from my friend, or that she asked him to answer it for her (say, she is driving). If I am pretty sure that the number is correct, these last explanations may be plausible. But suppose that the person answering maintains that he owns the phone with that number (perhaps, wishing to dispel my doubts, he emails me documents proving that this is so). In this case, the hypothesis that, after all, I got the number wrong becomes much more likely. Of course, there could be further alternative possibilities: maybe there is some conspiracy, involving secret espionage agencies and forged ownership documents. But this explanation is so convoluted and extraordinary that I will only be willing to consider it if I am really sure that the number is the right one (say, I have been phoning my friend every afternoon for the last ten years).

¹¹⁴ Similarly, if I want to check whether the glass door just in front of me is closed, I can extend my hand and see whether I encounter a solid surface. There might be odd circumstances in which my extended hand fails to encounter a solid surface even if the glass door in front is closed. But these possible impediments are so odd that I would only consider them in very special contexts (in light of very special evidence).

As this example shows, there is an interplay between the level of certainty of the agent's beliefs (i.e. how much rational support they have) and the kind of obstacles that count as relevant enough: in order to account for the failure of an action based on strongly trusted beliefs, it will be relevant to consider the possibility of obstacles whose relevance would be otherwise very low – obstacles that are unlikely to take place, given the evidence available, even if they cannot be discarded altogether. So, when physicists at CERN measured particles faster than light (something against their most entrenched beliefs), they were disposed to investigate the possibility of all sort of non-obvious mistakes and impediments (e.g., malfunctioning in routine components of the experimental set-up), before concluding that their beliefs about the speed of light and special relativity should be revised. If the beliefs called into question had been less resilient (and, in particular, less supported by practical success), scientists would probably have revised them without needing to check the absence of all these far-fetched impediments.

The practices of investigation and belief updating that I have sketched are defeasible all the way down. The conclusions reached in this sort of inquiry are always revisable – they can be amended in the light of new evidence. Every step in the process – including discarding or accepting the presence of obstacles – is carried out under the assumption that certain possibilities are not taking place, assumptions for which no positive reasons need to be provided (unless extraordinary evidence appears). Each of these possibilities may become relevant in some context, where they need to be examined explicitly, but they cannot be questioned all at the same time: there always needs to be some underlying background which remains, for the purposes of that particular inquiry, undoubted (all these points are familiar from discussion's dating back at least to Wittgenstein's *On Certainty*). So, for instance, in most cases scientists will not doubt their recent memory, or their being awake rather than dreaming.

7.6 Defeasibility as a source of inferential occasion-sensitivity

In this section I will discuss the way in which the relevance of possible defeaters (and how relevant a possible defeater needs to be in order to require explicit consideration)

depends on a variety of practical factors and features of the context in which the agent performs her actions.

The non-monotonic character of reasoning based on evidence offers a general account of the sensitivity of epistemic claims (in particular, knowledge claims) to contextual or pragmatic factors, which has been widely discussed in the recent literature (see, for instance, Stanley 2005; Fantl and McGrath 2007; MacFarlane 2005; DeRose 1992; Chrisman 2007; Williams 2001).

I have been arguing that, in non-monotonic reasoning, agents are not required to rule out explicitly each possible defeater, but only those whose possibility is sufficiently relevant – that is, an agent only needs to discard explicitly the possibility of a certain defeater if there are good enough reasons for taking the presence of such a defeater as a likely possibility, or at least a possibility worth considering. Now, how many possible defeaters – and with what degree of relevance – do agents need to rule out explicitly before counting as having sufficient reasons for endorsing some conclusion? Plausibly, the answer depends on different features of the context. In particular, practical stakes and costs will have an impact in fixing the threshold that determines how relevant defeating conditions need to be in order to require consideration. In low stakes situations, unlikely defeaters may remain unconsidered, whereas in a high stakes scenario those same defeaters (with the same likelihood) may need to be explicitly discarded – in order for the agent to count as having enough reasons for endorsing her conclusion. So, if the consequences of your eating nuts are trivial (you just do not like the flavor), the waiter's assurances that the food in the menu does not contain nuts may count as giving you enough reasons to eat it; by contrast, if you are very allergic to nuts, you may need to discard many possible defeaters before counting as having enough reasons to conclude that the food in the menu is free from nuts and is thereby something you can eat: perhaps you will need to check directly with the cook so as to make sure that the waiter is not mistaken; or maybe you will need to rule out the possibility that the ingredients employed became contaminated with nuts before entering the kitchen (and so on).

Thus, pragmatic and contextual factors – in particular what is at stake and the urgency and costs of making a decision – determine how relevant a defeating possibility has to be for it to require explicit consideration. In this way, whether a given set of premises counts as providing (enough) reasons for endorsing some conclusion – so that

the piece of reasoning from the premises to the conclusions is good –, may depend on pragmatic and contextual factors. One possible view is that these pragmatic factors shift epistemic standards – for instance, they may shift the threshold for some evidence to count as providing knowledge (MacFarlane 2005; DeRose 1992; Stanley 2005; Fantl and McGrath 2002, 2007).¹¹⁵ An alternative position is that epistemic standards (e.g. standards for knowledge) remain invariant (Reed 2010), but that the degree of epistemic support needed to be entitled to endorse a given conclusion varies with such pragmatic factors (see Brown 2008). So, according to this second position, it could be that in high stakes situations agents need to possess a degree of certainty close to full knowledge (i.e. they need to rule out almost all possible defeaters) in order to be entitled to draw certain conclusions, whereas in low stakes circumstances such levels of certainty are not demanded. I will not discuss here which of these options is preferable.

Besides setting the threshold of relevance for defeating possibilities to require consideration, the circumstances surrounding the occasion of acting may also contribute to determining *how relevant* the possibility of a given defeater is. Sometimes, a given defeating possibility will be likely because of the environmental conditions typical of the scenario in which the action is carried out. The typical environmental conditions of some scenario may give reasons to suspect that a certain defeater may be present, a possibility that in other scenario may remain unconsidered, as extremely unlikely (as seen in 7.4 when discussing fake-barn scenarios). Imagine, for instance, that you want to cook pasta in boiling water. In a standard Earth-surface scenario, pressures are typically around 1 atmosphere; thus, you will not need to consider the possibility that the pressure may be much lower than 1 atmosphere (something that would make water boil at a lower temperature, preventing the pasta from cooking). At least, you will not need to consider such a possibility unless extraordinary evidence made it relevant. By contrast, if you were in a scenario where very low pressures are not unusual (say, the surface of Mars), then you may be required to consider explicitly the possibility that pressure is low, so the temperature of boiling water is not enough to cook pasta: in those circumstances, such a possibility is relevant.

¹¹⁵ I will not discuss whether the relevant pragmatic factors pertain to the context of the subject to whom knowledge is attributed (Stanley 2005), the context of the attributor (DeRose 1992) or the context of the evaluator assessing the attribution (MacFarlane 2014).

Furthermore, as discussed in section 7.4, whether some defeating possibility is rationally expectable (and thereby relevant) depends on the background evidence of agents and on the information available in their context. Actually, I argued that a defeating condition may become relevant due to misleading information possessed by the agent (remember the example about the generally reliable encyclopedia with false information about the presence of venomous snakes in a certain island).

So, summing up, the relevance of the possibility of a given defeater, and how relevant the possibility of a defeater needs to be in order to merit consideration, depends on different practical factors and on features of the circumstances surrounding the occasion of acting – in particular, features of the context of the agent (and perhaps also of the context of those assessing the agent's actions). These situational factors include typical environmental conditions, the information available to the agent and how high stakes are. As a result, a defeater that may need to be explicitly ruled out on one occasion of acting may remain unconsidered on other occasions (even if defeaters are actually absent on both occasions).

It is interesting to note that, as a result, inferentialism becomes a form of context-sensitive semantics. In the sort of inferential semantics I have been presenting, the content of claims is determined by their role in practices of giving and asking for reasons. This includes (at least in Brandom's model, and I have followed him in this point) the role played by claims in defeasible pieces of reasoning – in particular, defeasible pieces of practical reasoning. But, I have just argued, whether a given possible defeater has to be explicitly considered by the premises of a piece of reasoning in order for it not to be enthymematic depends on features of the occasion of inference. Only defeaters whose possibility is relevant enough have to be explicitly discarded. And it is a context-sensitive matter i) how relevant a given possible defeater is and ii) how relevant a possible defeater needs to be in order to merit explicit consideration. Thus, what counts as good reasons for what depends on features of the occasion of inference: what inferential entailments are good is a context-sensitive issue. The same inference may count as good on some occasions and enthymematic on others (where further possible defeaters need

to be explicitly ruled out by additional premises). This introduces an occasion-sensitive dimension into the semantics of claims involved in defeasible inferences.¹¹⁶

I will conclude this chapter discussing where these epistemological considerations leave the pragmatist approach to representation.

7.7 Defeasibility and representation

The fact that practical reasoning is in general defeasible does not mean that the process of belief updating resulting from our evaluations of practical success does not answer to the way the world is. In this section I will examine the picture of representational practices emerging once defeasibility is taken into consideration.

On the view I am proposing, assessments of the success of an agent's action will lead to updating her cognitive views. Although conditions of success do not specify truth conditions for individual beliefs, beliefs are to be revised in systematic ways in view of the outcomes of actions. So, the success or failure of a given practical endeavor is followed by a pattern of revisions and adjustments in the agent's views. What specific adjustments are made does not depend only on the success or failure of that particular action, but also on how such success or failure is integrated in a wider web of practical interactions with the environment. In particular, whether a given belief is revised in light of the result of some action will depend on the outcomes of related actions and, crucially, on the context surrounding the agent's doings. Being a rational agent amounts to knowing what practical consequences may be expected from the accuracy of certain beliefs in a given context,

¹¹⁶ I will not discuss here in detail how this inferentialist occasion-sensitive semantics should be developed. One option is to argue that the content of claims changes across occasions of inference (the content expressed by assertions of the sentence 'The water is boiling' would differ in contexts where different possible defeaters are relevant). An alternative option is to see the content of claims as a function from collateral beliefs and occasions of inference (associated with a background of relevant defeaters) to inferential significance – remember that in chapter 3 I defined the inferential significance of a claim as the claims following from (and followed by) it. On this second view, the content of claims would be invariant across contexts of inference, even if the validity of inferential entailments involving such claims changes from context to context.

and also how these expectations are altered if some defeater is detected (for instance, in view of the outcomes of other actions).

Notice that the fact that detached truth conditions cannot be derived from success conditions does not make the content of individual beliefs indeterminate or lacking representational purport. In the inferentialist framework I am presenting, the content of claims is determined by their inferential connection with other claims, and also by their role in observational reports and in licensing actions. All these connections form a rich web that characterizes completely the content of claims (or, more modestly, a web that characterizes aspects of such contents which are central in explaining their deployment in reasoning and communication). Certainly, one of the results of the analysis of defeasibility developed above is that the shape of these inferential webs is sensitive to features of the context and of the circumstances surrounding the situation represented. I have argued that contextual factors determine, among other things, whether the presence of a certain obstacle has to be ruled out explicitly by one of the premises of the piece of reasoning licensing some action – or whether, on the contrary, such a premise does not need to be included in order to count as having good reasons for performing the action. For instance, for actions carried out in standard Earth-surface environments one does not typically need to discard explicitly the possibility that the atmospheric pressure is far from 1 atmosphere; this possibility, by contrast, may have to be considered in environments where pressure is often much lower or higher than 1 atmosphere (say, a special laboratory, or the surface of the Moon). This contextual dependence, however, does not make contents indeterminate or somehow defective: it only introduces the requirement that speakers grasping the contents of claims must show enough sensitivity to these contextual features, and to the way they affect the inferential behavior of such claims.

Assuming that agents' evaluations of their practical success are sensitive – at least to some extent – to the way things are in the world, the pragmatist model I am exploring gives an account of how our cognitive views are shaped by the world around us; in other words, this model would explain how the world has a say in determining the correctness of some of the moves of discursive practices. Our representational views respond non-arbitrarily to how the world is – as revealed to us through our evaluations of practical success. A belief has the representational purport it has by virtue of being revisable in

certain ways – in coordination with other beliefs – in the face of feedback from practical interactions with the world. This process of revision of our views in light of their practical outcomes is not piecemeal (that is, the evaluation of the outcomes of a certain action is not linked to the updating of a single, isolated belief), but this does not mean that such a process is not constrained by how things are in the world. True, the resulting picture presents the relation between our representational practices and the world as holistic, movable and subject to constant revision and amendment (rather than fastened to unshakeable foundations). But this seems to be the nature of our representational grasp of the environment. We keep updating our views in open-ended ways so that they best adjust to, and guide our interactions with the world.

7.8 Conclusions

In this chapter I have discussed how the issue of ignored obstacles can be addressed within the inferentialist framework introduced in chapter 4. The possibility of unconsidered obstacles for the success of actions is reflected by the non-monotonic character of the relevant pieces of practical reasoning – or, analogously, of the corresponding means-end instrumental theoretical inferences. Such pieces of reasoning may become bad if new premises (i.e. defeaters) are introduced, even if those new premises are not incompatible with the old ones.

Thus, the possibility of unconsidered obstacle can be accounted for naturally in the inferentialist model presented in chapter 4, by treating practical reasoning as defeasible. However, this defeasibility thwarts the project of deriving the truth conditions of individual beliefs from the success-conditions of the actions following from them: success-conditions are inextricably merged with anti-defeater conditions (a problem that was already discussed in chapter 5 in relation to success semantics). I have argued, nonetheless, that this does not mean that beliefs cannot be revised and updated as a result of the evaluation of the outcomes of ensuing actions. In general, agents may be attributed (defeasible) entitlement to endorse some conclusion even if they have not discarded explicitly all possible defeaters: it is enough if they do not possess reasons for taking the possible presence of such defeaters into consideration. Moreover, when an

Chapter 7

action fails, not all possible sources of such failure will be regarded as equally relevant: in a given context – given the rest of the evidence possessed by the agent, and the outcomes of related actions –, only certain causes of failure will be contemplated. In particular, many possible defeaters will remain unconsidered, unless special evidence appears providing reasons for thinking that such defeaters might have obtained.

The upshot of the discussion in this chapter is that, even if the correctness of our representational views answer to the way the world is, this responsiveness is not piecemeal but holistic: it is complex webs of beliefs and assumptions that are adjusted in response to the outcomes of our dealings with the environment (although some parts of these webs tend to be more easily revised).

According to the pragmatist approach I have been presenting, discursive practices involve friction with the world because agents take their evaluations of practical success to determine the correctness of certain moves within the practice, and I have been assuming that such evaluations have some sensitivity to the way things are in the world. In the next chapter I will discuss this assumption. This discussion will be related to the second problem faced by success semantics, namely, the determination of the conditions of success of actions.

Chapter 8: Representational Practices and Environmental Constraints

8.1 Introduction

The main thesis of the pragmatist approach to intentionality explored here is that discursive practices count as having a representational dimension – they count as answering to the way the world is – because they involve interactions with the environment, and the results of those interactions depend on the way the world is. More specifically, the pragmatist proposal I am developing is that the agents' assessment of the outcomes of actions will determine what is to be held as correct within the practice.

It is essential for this project, therefore, to examine how the world shapes the outcomes of our practical agency. In particular, it is necessary to discuss in which way the assessment of the outcomes of our actions can be said to answer to the way the world is.

In chapter 6 I discussed the possibility of accounting for the conditions of success of actions in terms of biological goals (Papineau 1986; Millikan 1984). I argued that neither evolutionary, nor organizational views of biological goals offered suitable grounds for accounting for intentional agency. My main criticism was that the goals identified by these approaches can diverge radically from the behavior of the agent to whom the goals are attributed.

In this chapter I will discuss how success conditions can be best accounted for in the pragmatist framework I am considering. I will propose a view according to which the goals of an agent are determined by the monitoring and regulating behavior of the agent herself – or, in general, of agents belonging to the same community. In this approach, agents can be said to be subject to goals because they treat performances as succeeding or failing to satisfy such goals. In the second part of the chapter I will discuss whether this sort of pragmatist approach leaves room for objective representation – that is, for a notion of correctness that is determined by the way the world is, rather than merely by how it is taken to be.

I will first argue that objective representation requires that there is a distinction between an action being *actually* successful and an action merely being evaluated as successful by some agent. This distinction only makes sense if, at least in principle, the evaluating responses of agents could be mistaken – there must be room for evaluative error. I will claim that the best way of allowing for the possibility of evaluative error is by appealing to alternative evaluative perspectives: the evaluations made by some agent may be evaluated as mistaken from some alternative perspective (say, the perspective of other agents). In objective representational practices instituted by fallible agents like us, it has to be an open possibility that any evaluative perspective may turn out to be mistaken (since in such practices correctness is determined by the way the world is and not by the attitudes of agents). This open-ended possibility of error can only arise if evaluations of success are made in a defeasible, revisable way, so that any evaluation could be eventually revealed as mistaken. By putting forward their evaluations of success only as *defeasibly* correct, fallible agents may see themselves as entitled to making such evaluations while taking them to be subject to objective standards of correctness.

Revisability, thus, is a necessary condition for fallible agents to engage in objective representational practices. However, revisability is clearly not sufficient for the emergence of objective representation. The fact that every evaluation may be revealed as mistaken or incorrect from the viewpoint of some alternative perspective does not mean that any of such evaluative perspectives is answering to the way the way the world is – the mere existence of clashing perspectives does not need to make the practice sensitive to how things are in the world. The evaluative attitudes of agents, besides being susceptible of error, must introduce friction with the world in some way. In the last part of the chapter, I will argue that it is a presupposition of our representational practices that the evaluative attitudes of agents do track, by and large, relevant environmental features. If our evaluations were not sensitive, to a large extent, to how things are in the world, it would not make sense to say that the practices we engage are representational. This does not mean, however, that when agents are in a scenario in which their evaluative attitudes do manage to track environmental features, such agents should not count as engaging in fully objective representational practices – even if they would not be able to distinguish, from inside the practice, some alternative scenarios in which their attitude fail to track features of the environment.

My conclusion will be that the pragmatist view endorsed here manages to offer a suitable characterization of the kind of objectivity exhibited by human discursive practices. Fallible agents like us can be said to engage in objective representational practices as long as their evaluative attitudes are, first, taken to be defeasible and, second, manage to track – by and large – relevant environmental features.

The structure of the chapter is as follows. In section 8.2 I introduce my proposal and relate it to the discussion about self-regulation and biological goals in section 6.5. Then, in section 8.3, I argue that the view of goals I put forward can make room for the possibility of error and may account for the objective dimension of representational practices. In section 8.4 I discuss what distinguishes objective and relativist discursive practices. Finally, in section 8.5 I examine in what sense the practices I have been described can be said to be linked to the world.

8.2 Self-regulation and goal-directed behavior

In chapter 6 I tried to find the best possible support for the claim that either natural selection (according to evolutionary approaches) or self-maintenance (according to organizational views) introduces an observer-independent evaluative standard. My strategy was to interpret environmental pressures as selective sanctions. Interactions with the environment would tend to reinforce certain traits and behaviors, and such reinforcement would – according to the argument proposed – reveal an underlying observer-independent evaluative standard.

A first problem with this strategy is that talk of selection, sanctions or evaluations in relation to environmental pressures borders the metaphorical. In paradigmatic cases, selective sanctions are issued by agents – rather than natural forces.

A second problem, as explained in chapter 6, is that the goals identified by these approaches can diverge systematically from the sort of goals suggested by the behavior of individuals. This is problematic because such behavioral cues play a key role in our intuitive ascriptions of intentional agency to individuals. Thus, it is not clear that the sort of biological goals characterized by these approaches may ground a suitable account of intentionality.

In this section I will present and defend an alternative proposal, which makes it possible to overcome the two problems just mentioned. Rather than thinking of the environment as a source of sanctions and evaluations, the idea is to pay attention to how goals are manifested through the evaluative attitudes of the agents subject to such goals. This proposal is in line with pragmatist accounts of normativity, such as those found in Beisecker (1999), Haugeland (1998), Kukla and Lance (2014) and Brandom (1994). I will try to develop this sort of approach in a way that manages to make room for objective representational practices – dealing with the difficulties I identified in chapters 5 and 6.

In the approach I will explore, the focus is on the way agents themselves sanction and regulate their behavior – and that of other agents in their community. On this view, goals arise because agents treat their own performances as successful and failed. Something counts as a goal of certain agents if those agents evaluate their behavior as successful or failed in relation to the achievement of that goal – where the evaluation may merely consist in some relevant form of self-regulation or sanction: agents would be disposed to reinforce and promote performances or states counting as successful, and to inhibit or suppress failed ones.

Note that this proposal differs from the view that goals are attributed by external observers: here, the evaluations giving rise to the goal are made by the agent herself – or, at any rate, from within her community.

This alternative outlook is free from the two difficulties presented above. First, sanctions are issued here by agents, rather than impersonal forces. Behaviors are selected and evaluated by agents. When it is an individual organism that monitors and regulates its own behavior – and maybe also the behavior of others –, it becomes less forced to describe the situation in terms of sanctions and evaluations.

The second problem – that is, the decoupling between the agent's behavior and the goals ascribed – is automatically avoided in the present approach, because goals are identified precisely through their expression in the behavior of agents (in particular, their regulatory behavior). So, the possibility of radical clashes between the goals identified and the attitudes of the agents subject to them disappears, or at least is severely limited. This is good news since, remember, in the sort of pragmatist framework I am presenting,

the success or failure of some action will only lead to updating the views of the agent insofar as she assesses her action as successful or failed.

This tight connection between the goals attributed and the attitudes of agents brings, however, its own problems. The most obvious one is that it seems that this approach cannot account for objectivity, which I have been considering a fundamental ingredient of representational practices. As I understand the notion, objectivity requires that what is correct is not wholly determined by what is taken to be correct, but rather by the way the world is. I have been arguing that the correctness of our representational views is constrained by the success or failure of their practical consequences, and that such success or failure depends on the way things are in the world – not only on the attitudes of agents. However, the account of goals I am presenting now seems to maintain that goals are determined by the attitudes of agents: being successful would just amount to being treated as successful. Thus, there does not seem to be a difference between being actually successful (given the way the world is) and merely being taken to be successful. Objectivity would be lost.

Another way of putting the worry is to notice that the evaluations of agents would seem to be infallible: successful behaviors or states would be determined by what agents evaluate as successful. Thus, there is no room for error in the agent's evaluations. If this is so, it is not clear that discursive practices answer to the world in any relevant sense – if agents evaluate something as successful or failed, it will be so, regardless of how things are in the world.

Note that this lack of objectivity is not problematic when accounting for many evaluative standards. There are many socially instituted standards which, actually, answer only to the attitudes of the relevant agents – so that the world cannot make such attitudes mistaken. Think of the status of being cool in a high school environment, or being well-mannered in high society meetings. Arguably being 'cool' in a high school context, just amounts to being regarded as cool by your relevant peers: if one is not considered cool enough, one is not cool enough – there is not further fact of the matter. In the same way, if eating with your hands is taken as good manners in some society, then eating with your hands is being well-mannered in that society.

However, this sort of goal (being cool or showing good manners) is not suitable for grounding an account of intentionality. Intentionality seems to require goals whose achievement depends on how things are in the world, rather than only on the evaluations made by agents; in other words, being actually successful must not just consist in being evaluated as successful by agents, but in the fact that the world is in certain success-conferring ways (a lion can be said to be looking for food only if its goal is satisfied by actually finding food, as opposed to merely considering that food was found).

It must be stressed that evaluative standards do not become objective just because the practices they emerge from are world-involving – that is, include interactions with the world (see Kukla and Lance 2014; also Brandom 2008: lecture 6). For example, the practice of dining clearly involves dealings with the physical world: the way the world is constrains human eating practices. Nevertheless, it seems that whether it is good manners to eat with your hands depends only on the attitudes of the relevant agents; therefore, the standards of good manners for eating are not objective – even if the practices they are embedded in are world-involving.

These difficulties with the objective dimension of representation pose, in my opinion, the hardest challenge for the account of goals I am presenting – and in general for the pragmatist project developed here. In the rest of the chapter I will try to argue that a satisfactory reply may be given – at least, a reply that captures those aspects of objective representation characteristic of human practices. I will first show how the possibility of error of evaluation may be introduced in the sort of practices I am describing. Then, I will discuss whether the evaluative standards arising in such practices can be said to answer to the world in a relevant sense – that is, whether the world, after all, gets to have a saying about the correctness of the moves of the practice.

8.3 Making room for error

8.3.1 Error and alternative perspectives

In a straightforward reading of the proposal I have introduced above, the evaluative attitudes of agents (their assessment of the success of a performance) are incorrigible, since they establish the standard against which performances are to be evaluated. Success would be wholly determined by evaluative attitudes, so there is not a distinction between being actually successful (actually achieving some goal) and being merely regarded as successful.

This incorrigibility seems to be at odds with the possibility of objectivity – since objectivity seems to demand that it is a possibility, at least in principle, that our subjective attitudes get things wrong (i.e. that our attitudes are mistaken about the actual achievement or failure of some goal).

The best way of allowing for error in the evaluations of agents, I believe, is to take into account multiple evaluative perspectives (for this sort of approach, see Brandom 1994; Haugeland 1998; Beisecker 1999; Kukla and Lance 2014). These different perspectives may clash, giving rise to the possibility of appreciating mistakes: the evaluations made from one perspective may be regarded as incorrect from an alternative perspective. The notion of evaluative error will be articulated to a greater extent as the web of potentially conflicting perspectives gets more complex.

As I see it, the most natural way to go is to associate this multiplicity of evaluative perspectives with a multiplicity of agents.¹¹⁷ Different agents would evaluate the success of a given action from different – potentially diverging – points of view. But perhaps there can also be diverging perspectives belonging to a same individual: they may derive, for instance, from different sensory modalities, from different ways of evaluating related goals or they may be evaluations made by the same individual at different times.¹¹⁸ In

¹¹⁷ As explained in chapter 2, Brandom (1994: ch. 8) argues that this sort of social perspectival articulation is a necessary condition for objectivity.

¹¹⁸ Brandom characterizes the relation between different temporal perspectives within the same individual as a form of social relation: "*if* creatures *can* take up the different perspective to time-slices of themselves,

general, it seems that there could be relations among different evaluative responses of a single individual, so that some such responses are taken to be incompatible (at least one of them must be mistaken). One way in which this can happen is if there are alternative paths for evaluating a given performance. Disagreement in the evaluations reached through these alternative paths may be taken to indicate that some of those evaluations were mistaken. So, for instance one can assess whether the cake is burnt by looking at it and also by tasting it.

Likewise, the evaluation of the outcomes of a given action will serve as input for further evaluation-performance feedback loops, which will involve further evaluations. These interweaved loops may yield clashing evaluations, which may lead the agent to regard some of those evaluations as mistaken. For example, imagine that I assess that I have succeeded in mixing the right amount of baker's powder in my cake, and proceed to put it in the oven. However, when I take it out I see that I have failed to obtain a spongy cake – what I find is a flat sort of pancake. This may give me reasons to think that, after all, the amount of baking powder was not right.

This sort of interrelation among different evaluations will be more developed in agents with the sort of capacities characteristic of full-fledged mastery of concepts; that is, agents that are able to track the consequences and implications of a certain evaluation – in particular, what further evaluations will be incompatible with that first one (so, an agent possessing the concept “baking powder” may be able to think that a cake with enough baking powder will be spongy, rather than flat). In agents with this sort of capabilities, the assessment of any given action will have repercussions for wide array of other evaluation-performance loops: therefore, there will be a complex and extended web of evaluations that may clash with each other, giving rise to the possibility of treating some of those evaluations as mistaken. Think, for instance, of the scientist who extracts far-reaching consequences from the assessment of the outcomes of a particular

then the relation among those time-slices is *social* in the sense that it must admit of the distinction of perspectives between the attitude of attributing a commitment (or other normative status) and the attitude of acknowledging it" (2010: 299). But, of course, an account of normativity in terms of intrapersonal time-perspectives will count as individualistic (see Gibbard 1996; Blackburn 1984).

experiment – consequences affecting the correctness of several theoretical claims, and also the status of other experimental results.¹¹⁹

I have granted that there may be alternative assessing points of view within a single individual; however, such multiplicity of alternative perspectives is crucially enhanced by the social context in which representational practices are embedded. So, the success of the actions of an agent is not only evaluated by the agent herself, but also by other speakers – who will assess accordingly the agent's entitlement to further actions. Moreover, the other speakers may revise their own views, and their own practical plans, in the light of the outcomes of the actions of an agent.

In this way, evaluation-performance cycles are integrated in a complex matrix of social interactions, which introduces a host of additional points of view – of additional assessing perspectives. As a consequence, room is created for the possibility of the incorrectness of any individual perspective (something that, in the absence of a social context, can only be done to a limited extent). It may always happen that the views held by any individual, or group of individuals, are evaluated as inaccurate from the perspective of another agent (see Brandom 1994: ch. 8).¹²⁰ The assessments of an

¹¹⁹A good example was the surprising measure in CERN of particles faster than light. Given the unexpected nature of such measure, several related experiments were carried out, and it was finally concluded that the first experiment had been *incorrectly assessed* as a successful acceleration of a particle beyond the speed of light.

¹²⁰ It may be argued that, once the evaluative perspectives of other agents are taken into account in order to determine the success of an individual's actions, there reappears the risk of decoupling the attribution of goals from the behavior of the individual to whom the goal is attributed – remember that in section 7.3 I claimed that this was a problem for evolutionary accounts, in which the goals ascribed to individuals could diverge radically from the individual's goal-seeking behavioral marks (as happened in the example of the homosexual rabbit).

There may be cases in which the social network of evaluative perspectives ascribes goals to an individual overriding her own evaluative behavior (e.g. when the social environment of a teenager assesses her actions in relation to the goal of, say, becoming a successful business woman, even if she clearly shows a preference for a career as a mathematician); however, I think that this is not an inescapable consequence of the present proposal. Discursive practices may be such that, in general, the evaluative behavior of an individual is conferred special authority in fixing her own goals – even if, once those goals are fixed, other individuals may disagree with the agent about whether those goals were actually achieved. So, it is up to me to decide whether I want to cook a carrot cake or an omelet; but, having chosen the cake as my goal, it

individual agent will have to pass through the feedback loops carried out by that agent, and also face the assessing stances of other speakers – and pass the tests posed by their interconnected evaluation-performance loops. Thus, social representational practices have a rich structure, layered with a multiplicity of crossing perspectives, in which a strong notion of error finds its place.

It has to be clear that I am not defending the view that communal assessments are necessarily more reliable than those of isolated individuals.¹²¹ My claim is only that in that sort of social contexts, the notion of error gets to be articulated to a greater extent – in virtue of the many alternative perspectives to be contrasted.

8.3.2 Objectivity and defeasibility

The existence of alternative assessing perspectives makes room for some form of error; but it does not automatically bring in an objective notion of correctness. For, when *all* alternative perspectives agree, it seems that there is no way in which such agreeing perspectives may count as mistaken – there are no further perspectives which may clash with the consensus view. It seems that objectivity would be substituted by collective accord, be it accord among intra-individual perspectives or among inter-personal ones (see McDowell 1998 for a critical appraisal of this substitution; also Wright 2001).

In order to get objective practices, it is required that no evaluation is actually correct merely by virtue of being made from a specific perspective – not even if it is the global perspective formed by the aggregation of all available perspectives. There must be always a distinction between being actually correct and being regarded as correct from

is not up to me to decide whether I succeeded in cooking it: even if I think I did, others may properly challenge my assessment (perhaps I did not taste it so I did not realize that the thing I produced is nothing like a carrot cake).

¹²¹ Although in many cases communities will have access to epistemic resources beyond the reach of isolated individuals, and as consequence they will be better filters of error.

some perspective (Brandom 1994: ch. 8 discusses in detail this requirement¹²²). Any perspective must be, in principle, challengeable.

Thus, it seems that a necessary feature of practices giving rise to an objective notion of correctness is defeasibility: any evaluation could in principle come out to be wrong – no matter from which perspective it is made. In a defeasible assessing practice, agents would always be open to the possibility of there being an alternative perspective that reveals their views as actually mistaken; thus, a strong distinction between being correct and being regarded as correct is preserved. By examining the implications of currently endorsed assessing attitudes and looking for incoherence within the network of intertwined evaluations, some views will be revised and treated as incorrect. As new evaluative attitudes are endorsed, the possibility of incompatibilities with previous evaluations arising remains indefinitely open: every attitude is, in this way, susceptible of revision and correction. It is for this reason that evaluative attitudes are endorsed in a defeasible way: if suitable challenges arise, such an endorsement may be revoked.

Full-fledged defeasibility, in turn, seems to require that the number of alternative assessing perspectives is open-ended. No matter how many perspectives agree on some evaluation, there should always be the possibility that there appears a further clashing evaluative perspective. Only this possibility ensures that all evaluations are revisable.

Again, this sort of open-ended web of assessing perspectives can in principle arise both at the social and at the individual level. At the individual level, I have argued that a given evaluation may be challenged by evaluations of the outcomes of further actions suitably related to the first one (e.g. my success in cooking the pasta is related to my success in heating up the water in the pot). As long as the agent has the ability to track far reaching implications of a given action, it seems that there will always be the possibility that she may come to yet another evaluation that clashes with her previous views.

¹²² In chapter 3.7, I discussed Brandom's claim that this objectivistic requirement is met by the sort of social practices he describes; Brandom's views on this point are notoriously opaque (for critical discussion see Rödl 2010; Hattiangadi 2003; Glüer and Wikforss 2008; Grönert 2008; Shapiro 2004; Wanderer 2008: 78-92; Laurier 2005; Rosen 1997).

So, the story above offers no a priori arguments showing that sophisticated versions of intentionality can only take place in social contexts. Perhaps, the different assessing perspectives attached to a same individual are enough to give rise to relevant kinds of objectivity. I will leave this question open. My impression, however, is that the web of interacting points of view supplied by social scenarios is far richer than anything accessible for an isolated individual, and that therefore objectivity may reach a fuller development in social contexts (it could even be that a recognizable notion of objectivity may only emerge in such contexts).¹²³

To recapitulate, in this section I have argued that in order for there to be a distinction between an action being actually successful (because of the way the world is) and an action merely being evaluated as successful from some (fallible) assessing perspective, it is required that every perspective in the practice is seen as in principle revisable. This distinction is essential for objectivity, and therefore the revisable, defeasible nature of evaluating perspectives is a *necessary condition* for the emergence of objective representational practices – at least if such practices are instituted by fallible agents.¹²⁴

¹²³ Moreover, social scenarios with potentially clashing evaluative perspectives are the natural context for the emergence of justificatory practices. In such practices, individuals do not only treat performances as correct or incorrect, but they may also try to defend the correctness of their performances from the challenges made by other evaluators (e.g. by showing how its correctness is related to the correctness of other performances, about which the challenger is in agreement with the agent challenged). Being able to engage in these practices underpins – in accordance with the inferentialist framework presented in chapter 3 – the acquisition of conceptual abilities. Also, this sort of justificatory practices is, arguably, the door to a distinctive kind of awareness which characterizes the way in which rational beings represent the world and act in it – as opposed to the skillful navigation of the environment typical of non-rational creatures. (For a discussion about this different kinds of intentional agency, see the essays in Schear (2013)).

¹²⁴ Perhaps there could be an objective representational practice instituted by infallible agents whose evaluations of practical success are not revisable (they are always correct). Given that human agents are certainly fallible and I am interested in examining human representational practices, I won't further explore this possibility.

8.3.3 Objective representation and fallibility

One possible worry about an objective, perspective-independent notion of evaluative correctness is that it may seem that individuals within actual human practices will never be in a position to attribute such a demanding form of correctness. All perspectives in the practice are fallible, so it can be argued that individuals will never be able to tell whether a given evaluation is *actually* correct (at most, they can tell whether they *think* that it is correct). Is objective correctness an ideal that finds no place in human practices? Perhaps, when members of a practice assess normatively each other, they do not do so in relation to an objective notion of correctness, but to less demanding standards – that is, standards such that fallible individuals can tell whether they are actually satisfied (for instance, the standard of having compelling evidential support).

In this section I will offer reasons for resisting this conclusion. I will argue that it is possible to find a place in human practices for attributions of objective correctness, as long as such attributions are seen as defeasible – as open to open-ended revision.

The notion of defeasibility makes it possible for objective evaluative standards – standards that are independent of perspectives – to emerge in human practices, even if no agent or group of agents can be said to appraise such standards infallibly. The fact that every perspective is susceptible of error does not mean that no evaluation within the practice can be counted as (objectively) correct – that objective correctness is beyond the reach of the perspectives of the participants of the practice. Evaluations made from fallible perspectives can be put forward as objectively correct, *albeit in a defeasible spirit*: it always remains an open possibility that further challenges appear showing that the evaluation was actually incorrect (in light of new evidence, of further incompatible evaluations). When it turns out that a given evaluation was actually incorrect, former attributions of correctness will be revoked. But, in the absence of reasonable challenges, it will be possible to grant correctness to evaluations made from perspectives within the practice – although every evaluation, including those made from the point of view of the whole community, is in principle revisable.

Taking defeasibility seriously, thus, allows one to see human representational practices as making room for a strong, objective notion of discursive entitlement (i.e. correctness) – a sort of entitlement that may not be possessed by agents making false

claims. Remember that when developing my proposal, I have assumed that one could find a place for this stringent notion of entitlement in human discursive practices (see chapter 3 and 4). My aim has been to show that the correctness of moves in discursive practices is determined by the way the world is.

Interestingly, Brandom – backing down on his objectivist aspirations – does not seem willing to take on board such a stringent form of entitlement in his reconstruction of the pragmatics of representational practices (Brandom 2000, 1995). Brandom’s motivation for rejecting this strong notion of entitlement is that he thinks that it is not clear that such a notion of entitlement could play any role in our actual practices: agents would never be in a position to attribute or to claim such entitlements. So, he argues that:

[...] the conditional

7. “If I am entitled to the claim that the swatch is red, then the swatch is red”

is not one that ought to be endorsed as correct in the sense of commitment-preserving, at least for any notion of entitlement that humans can secure regarding empirical matters of fact. It is, after all, an instance of the very implausible schema

8. “If *S* is entitled to the claim that the swatch is red, then the swatch is red.”

(Brandom 2000: 201)

What I am arguing, precisely, is that Brandom’s conditionals 7 and 8 are both true – despite Brandom’s claims to the contrary. If an agent is actually entitled to asserting some claim *that p*, then it is the case *that p* (i.e. it is true *that p*). Entitlement, as I am understanding the notion, requires truth – a similar point, using a different terminology, can be made by claiming that actual justification implies truth (pace Brandom 1995).¹²⁵ This is compatible with accepting that an agent may be blameless in doing something she

¹²⁵ See Littlejohn (2012) for a defense of the claim that epistemic justification entails truth.

was not actually entitled to do (perhaps she had what appeared to be perfectly good evidence).

I think that Brandom's project of characterizing objective representational practices is ill-fated unless one makes room for a strong notion of entitlement – which bars entitlement to false claims. Brandom (2000: 196-204; 1994: ch. 8) makes much of the point that objectivity requires distinguishing between the correctness of a claim and its being endorsed as correct from some perspective. Thus, the truth of a claim does not depend on whether some agent or group of agents is in a position to vindicate their entitlement to endorsing it: in this sense, truth does not entail (possession of) entitlement.¹²⁶ So far, I am gladly on board; but Brandom goes on too quickly to accept the reverse direction of entailment and argue that entitlement does not imply truth – so that an agent's being properly entitled to endorsing a claim does not guarantee that the claim is actually true. This move is, I think, highly problematic for someone like Brandom, who wants to credit representational practices with a strong enough objective nature. The difficulties are especially burdensome given the spare normative materials Brandom's theory avails itself of. Remember that the basic normative notions in Brandom's proposal are commitment and entitlement. A discursive move (e.g. an assertion) is only inappropriate or incorrect, in Brandom's system, insofar as it is not properly entitled. Thus, if one allows for proper entitlement for false claims, there is no sense in which such false claims are evaluable as inappropriate or incorrect.¹²⁷ And, if the correctness of a

¹²⁶ The question of whether truth implies entitlement is more subtle than what I have made it seem. Perhaps there is a sense in which the truth of the claim makes it so that there are reasons for accepting it. Being this as it may, I will commit myself to the minimal view that the truth of a claim does not imply that agents *have* reasons to accept it (for the contrast between there being reasons/warrants for something and agents having reasons/warrants for doing something, see Whiting 2013). This is what I mean when I say that the truth of a claim does not entail possession of entitlement.

¹²⁷ One may devise an alternative theory in which one normative notion – say, entitlement – tracks the responsibility or blameworthiness of agents and another normative notion – for instance, appropriateness – tracks the objective correctness of discursive moves (so that there can be entitled but inappropriate moves). But this is not the way in which Brandom constructs his theory and, anyway, this notion of entitlement could not do the work that it currently does in Brandom's proposal (e.g. agents could be entitled, in this modified sense, to incompatible claims, insofar as an agent can endorse incompatible claims blamelessly – perhaps the incompatibility is subtle, and the agent was relying on convincing but misleading

discursive move is not determined by how things are in the world, it is dubious that such a move can be said to be representing the way the world is – at least, such a move would not be part of the sort of objective representational practices that I am trying to characterize.

As I have just explained, the best strategy for constructing an objective, stringent notion of entitlement out of the evaluative attitudes of agents is to treat such attitudes as essentially defeasible and open to revision – their entitlement can always be withdrawn in the face of new challenges. In this way, an objective notion of entitlement may have a place in actual human representational practices despite the fact that no evaluative attitude can be guaranteed to possess such entitlement.

It is important to note that I am not claiming that the defeasibility of the evaluative attitudes of agents is *sufficient* for the emergence of objective representation; all I am saying is that it is a *necessary* condition. Fallible agents can engage in objective representational practices *only if* their evaluating attitudes (in particular, their evaluations of practical success) are taken to be defeasible. In the next section, I will discuss the insufficiency of such defeasibility for securing objective representation: a practice in which all evaluative attitudes are in principle revisable may fail to count as a practice in which agents engage in objective representation of the world. I will present certain forms of relativism as an example. After that, in the final section of the chapter, I will discuss what further conditions need to be met by discursive practices in order to count as being in the business of representing the world objectively. More specifically, I will argue that representational practices require that evaluative attitudes are sensitive to the way the world is: evaluations of success have to track, by and large, relevant environmental features. When this further condition is met, I will claim, a minimal form of objective representation of the world can be said to arise.

evidence). Moreover, one would still have to be able to give a pragmatist account of the emergence of the objective notion of appropriateness in human representational practices.

8.4 Objectivity and relativism

In order for the practice to count as objectively representing how things are, it does not only have to be open-endedly revisable: the process of revision must introduce friction with the world. The progressive revision and correction of the evaluative attitudes of agents has to be, by and large, sensitive to the way the world is. Otherwise, it would seem that the practice cannot be distinguished from a relativist one.

In this section I will argue that relativist practices (at least when constructed in certain ways) can share most of the features of objectivist practices described above – in particular, their revisable nature. I will then suggest what further aspects of objective practices distinguish them from relativist ones. In the rest of the chapter, I will show how the role of practical interactions with world in representational, empirical discourse allows such a discourse to count as objective.

By relativism I will refer to the view that the truth of a given claim (as asserted by a certain speaker) depends on the perspective of evaluation (see MacFarlane 2014 for a detailed analysis of this form of relativism).¹²⁸ In particular, I will consider relativist practices in which truth is made relative to the attitudes or assessing responses of the evaluator occupying the context of assessment. For instance, in a simple relativist construal of gastronomic taste speech, the truth of claims about the tastiness of food (as assessed from a given context) would be relative to the gustatory attitudes of the evaluator: the claim “Apples are tasty” (as made by Mary) would be true from the perspective of James, who likes apples, and false from the perspective of Julia, who dislikes them.

Of course, this is a very rough and unrealistic model of speech about gastronomic taste. On a more sophisticated proposal, it could be that the truth of a claim attributing tastiness to apples (as evaluated from a given context) does not directly depend on the gustatory responses or attitudes of the evaluator, but perhaps on the responses she would have after careful reflection, or if she were fully informed and trained in the best

¹²⁸ This view is sometimes known as *radical relativism*, in order to differentiate it from *moderate relativism* (also known as non-indexical contextualism), which is the view that the truth of a given claim depends on the perspective of the speaker asserting it (see Kölbel 2015a, b for a review of different versions of relativism). I will focus on radical relativism.

cooking schools. In this way, the evaluator would be fallible about what taste standards are relevant in her context of assessment.

Moreover, as I discussed in chapter 3 (see note 48), there may be relativist practices in which truth is made relative to features of the context of assessment that are not determined at all by the attitudes of the evaluators in that context – so that there is room for objectivity, in the sense I am understanding it. The only claim I want to make here is that one could think of simple relativist practices, as the naïve model of taste speech sketched above, in which objectivity has no place.

In the inferentialist framework I have been working with, relativism would amount to the view that the normative status of a discursive move depends on the perspective of evaluation: in particular, whether a speaker is properly entitled to making a certain assertion would depend on the perspective of the evaluator. So, in the simple relativist model of gastronomic taste introduced above, it can be that from the perspective of James (he finds apples tasty), Mary is entitled to asserting “Apples are tasty”, whereas from the perspective of Julia – who dislikes apples –, Mary is not entitled to asserting such a thing.

In a relativist practice, thus, attitudes endorsed from any given perspective can in principle be criticized from an alternative perspective. In particular, MacFarlane (2014) explicitly describes relativist practices as susceptible of open-ended revision: there is always the possibility that I will come to occupy a further perspective from which my currently endorsed attitudes will be (properly) assessed as incorrect. The revisable nature of relativist discursive practices is captured by what MacFarlane calls the Retraction Rule (MacFarlane 2014: 108):

Retraction Rule. An agent in context c_2 is required to retract an (unretracted) assertion of p made at c_1 if p is not true as used at c_1 and assessed from c_2 .

According to the Retraction Rule, an assertion made – and held to be correct – in a first context (from a first perspective) ought to be retracted (taken back) whenever the speaker finds herself in a new context from which the assertion is evaluated as incorrect –

as lacking proper entitlement, in Brandom's terminology. Imagine that, in the past, I found that oranges were tastier than apples and I asserted that that was the case. By contrast, now I take apples to be tastier than oranges. According to MacFarlane's relativism, I ought to retract my past assertion, since, as seen from my present perspective, it was not properly entitled— even if it was properly evaluated as entitled from my previous perspective.

It seems that relativist practices will include as well the following Rejection Rule (MacFarlane considers this rule, but he thinks that Retraction is enough to distinguish relativism from other views, so he does not explore it in detail):

Rejection Rule. An agent in context c_2 is permitted to reject an assertion of p made at c_1 if p is not true as used at c_1 and assessed from c_2 .

(MacFarlane 2014: 110).

Rejection deals with interpersonal alternative perspectives. According to an inferentialist reading of Rejection, Julia would be permitted (i.e. entitled) to reject Mary's assertion of "Apples are tasty" – that is, to treat Mary as not entitled to making such an assertion. This is so because from Julia's perspective Mary asserted something false.

Relativist practices, therefore, are revisable: any given evaluative attitude may be challenged from alternative perspectives and agents may have to retract their former evaluations after coming to adopt a new assessing perspective. However, despite this revisable nature, relativist discourse about gastronomic taste, as characterized above, is clearly non-objective: what is properly assessed as correct from a given context of assessment depends on the (gustatory) attitudes of the evaluator occupying the context. The clash of assessing perspectives in relativist discourse about gastronomic taste (in the simple model considered here) does not aim at converging towards a more faithful depiction of the world, but at most at the mere coordination of the attitudes of the relevant agents.

Actually, from the point of view of the participants, a relativist practice may not be distinguishable from an objective one (that is, one in which truth or correctness does not depend on the assessing attitudes of evaluators). In both kinds of practices agents will assess perspectives incompatible with their own as incorrect or in error. Sure, when an agent in a relativist practice changes to a new perspective, she will regard her previous one as incorrect (e.g. she will retract the assertions she made from that perspective). But, from the vantage point of her current perspective, she may see her change of perspective as an improvement in her views, that is, as a transition from a perspective that is mistaken to another one that is not. The same will happen in objectivist practices: I recognize my past, abandoned perspectives as mistaken (otherwise, I would have no reasons to abandon them). Likewise, both in objectivist and (at least some) relativist practices agents leave the door open to the possibility that they will come to occupy a further perspective from which their current one will be revealed as mistaken.¹²⁹

What distinguishes, then, objective practices from the sort of relativist discourse I have described, if not their revisable nature? The difference, I submit, is that, in an objective representational practice, revisions must latch onto the way the world is – onto environmental trends and patterns. If the revision of the attitudes of agents is responsive to how things are in the world, there will be grounds for saying that such attitudes aspire to represent the way the world is, and that they are subject to objective standards of correctness. I have been arguing that representational discourse about empirical matters incorporates such friction with the world by virtue of involving practical interactions with

¹²⁹ One interesting question is whether a practice can continue to be relativist once its participants become aware that it was functioning in a relativist way – that is, once its participants realize that there is no sense in which any of two incompatible perspectives answers more accurately to how things objectively are. Arguably, when agents acquire such awareness, their practices will change; in particular, it is likely that they will become contextualist or subjectivist practices – that is, practices in which assertions are taken to specify how things are in relation to the speaker's perspective (e.g. in relation to her tastes). Once I realize that even if I am perfectly entitled to making some claim, other agents are, from their perspectives, also perfectly entitled to make seemingly incompatible claims, it seems that I will be inclined to renounce to treat any such perspective as reflecting more accurately than the others the way things objectively are.

I do not want to rule out the possibility that there are relativist practices that can survive this sort of awareness; I am only claiming that it is interesting to investigate whether this is so.

the environment; in particular, representational views are updated in accordance with the outcomes of such practical interactions.

In the last part of the chapter, I will discuss the way in which the world constrains the process of revising and updating the views of participants in representational practices – more specifically, the process of revising views as a result of the evaluation of the outcomes of practical dealings with the world. So far, I have tried to argue that error is possible in practices in which evaluative standards are set by the very evaluative attitudes of agents. Now, I will discuss in what sense such practices can be said to be hooked to the world, that is, to what extent the evaluations of agents pick up environmental patterns. Given that the gist of the whole pragmatist project presented here is that our practices count as representational because they are revised and updated in accordance with our evaluations of the outcomes of our actions, it is essential that, in some way, such evaluations depend on how things are in the world. When assessing attitudes are, first, revisable in open-ended ways (so that error always remains possible) and, second, sensitive to a large extent to environmental features, the resulting practice deserves to be seen as engaging in objective representation of the world – insofar as what is correct within the practice can be said to be determined by the way the world is.

8.5 The link with the world

In general, the results of our actions are not wholly controlled by us. An agent is not always able to bring about some practical outcome just by willing it: something else is required in order for such an outcome to obtain. It is natural to say that what is required is the collaboration of the world. Agents are not infallible about what their actions will bring. As Wittgenstein remarks: ‘I cannot bend the happenings of the world to my will’ (Wittgenstein 1979: 73).

This lack of control is exploited by the pragmatist approach to intentionality presented here. The basic idea is that agents regulate and evaluate their performances taking into account these beyond-control outcomes. The results of actions, as perceived by agents, do not always conform to the agents’ previous expectations or wishes. Thus, if

agents resort to their appreciation of such outcomes in order to evaluate actions, those evaluations will not depend only on their will or expectations.

Does the fact that agents cannot control certain outcomes mean that evaluations relying on those outcomes manage to capture some relevant features of the environment? Not necessarily. Imagine that Ann has a random itch, which does not depend on external stimulus. Ann cannot control her feeling itchy: the feeling comes and goes regardless of her wishing it to do so. However, the patterns of extinction and activation of the itch are not related to any specific environmental pattern. Similarly, it could be that someone hallucinates without being able to control what happens in the hallucination (it does not seem that we are able to control the outcomes of our actions in dreams).

A sceptic could argue that our whole experience of interacting with a not always yielding environment might be like a dream – in the sense that our evaluations of success and failure are not sensitive to any relevant environmental trend. True, I have appealed above to the alternative evaluative perspectives afforded by social contexts: but it could still be (according to the sceptic) that we live in some form of collective illusion or dream – for instance, we could live in the Matrix and be all coordinated brains in a vat. Insofar as one is willing to accept this sort of radical sceptical scenarios, it becomes a possibility that human practices fail to answer to the way the world is.

However, the fact that the participants of a practice would not be able to tell the difference if they were in some sceptic scenario (in which their evaluations were decoupled from the way the world is) does not mean that, when agents actually are not in such a sceptic scenario, their evaluations do not manage to track environmental features – in a way that makes such evaluations answer to the way things are in the environment (for a detailed discussion of this sort of view, see Williams 1991).

It is a precondition of representational practices, I think, that the evaluations of agents are to some extent constrained by the way the world is. If the world is completely decoupled from the agents' evaluative and regulatory behavior, it does not make sense to say that those agents are representing the world in any relevant way. Engagement in representational practices, thus, requires that the evaluative responses of agents are, by and large, suitably hooked to the world. When this requisite is met, there are no reasons

not to count the resulting practices as answering to the way the world is – so that representational correctness within the practice depends on how things are in the environment represented.

I will assume that our evaluations of success and failure (e.g. success in not being burnt when drinking tea) tend to capture relevant features of the world (e.g. being or not being actually burnt). Of course, this is the most plausible assumption if one follows current science in considering human beings as biological organisms trying to adapt to their environment. Biological organisms have been selected so as to self-preserve their organization and proliferate. As a result, their self-monitoring and self-regulatory behaviors will track down, by and large, features of the environment that are relevant for self-maintenance and reproduction. If this were not so, organisms would tend not to spread their genes.

So, the standards according to which actions are assessed are ultimately set by the evaluative attitudes of agents. It is those evaluations that make something count as successful or failed. However, in order to assess their actions, agents defer to the way the world is: agents look at the world in order to evaluate the success of our actions – and I am assuming that the resulting evaluations tend to be sensitive, in general, to relevant environmental features.

Thus, it is a constitutive feature of representational practices that the evaluation of actions is to be determined in accordance with the world, as manifested to agents in their practical engagement with the environment. If the world can be said to have any evaluative authority over our practices, it is because we *delegate* some such authority to it: it is, therefore, only authority by deferral.¹³⁰ Once it is agreed that evaluations are to be fixed in this way, though, it stops being up to agents whether some action is to count as successful or failed (remembers that agents could not control at will their responsive

¹³⁰ Rorty is typically associated with the view that normative authority always emanates from social practices (see Levine 2010). This sort of view was already put forward by Nietzsche: ‘Whatever has value in our world now does not have value in itself, according to its nature – nature is always value-less, but has been given value at some time, as a present – and it was we who gave and bestowed it.’ (1974: 242).

My claim is that Rorty's insight may be developed so as to leave room for the existence of objective representational practices (basically, this is the aim of Brandom's project (see his 1994)).

sensitiveness to the outcomes of their actions).¹³¹ Agents confer evaluative powers, so to speak, to the world. In this sense, it can be said that representational practices answer to the way the world is. By constraining the evaluative responses of agents, the world constrains the representational practices those agents are engaged in.

All this should not be interpreted as meaning that agents attribute infallible evaluative authority to certain moves in the practice – those related to their perceptions of practical outcomes. As explained in the previous section, objectivity requires that there is a distinction between a goal being actually achieved and a goal being merely evaluated as achieved. This distinction can only emerge in a context in which it makes sense to think of an evaluation as mistaken. I have argued that such a context appears when multiple alternative evaluating perspectives come into play. In these scenarios, any evaluation may be regarded as mistaken from a contrasting perspective. However, the structure of the practice will tend to make certain perspectives appear, under certain conditions, as less prone to be doubted – at least until further challenges arise. Also, the nature of human representational practices (their adaptive roots) will guarantee that, at least in some circumstances, such trustworthy perspectives are sufficiently successful in capturing relevant environmental patterns. When this is not the case, the self-maintenance and proliferation of those agents becomes seriously threatened.

Of course, one could think of a community whose members choose to make seemingly random evaluations – evaluations which fail to track down any relevant environmental pattern (e.g. they do not distinguish the results of eating edible and indigestible berries). One first thing to say is that, insofar as their evaluations are completely unrelated to environmental features, the members of this community would not be playing a representational game – they would be engaging in a different sort of practice. Moreover, all chances are that this community would quickly disappear – it would not be able to achieve self-maintenance in a competitive and selective

¹³¹ Think of a governing council that decides to trust the advice of experts about certain technical issues. The members of the council agree that, whatever the experts say, it shall be done. The council remains the source of authority, but it transfers some of such authority to the experts. If one of the members of the council acted against the experts' advice, she would be regarded as doing something inappropriate, since the council had agreed that its members were supposed to follow the experts' judgment in these issues.

environment. This sort of communal practice is, perhaps, conceptually possible, but, certainly, biologically implausible.

Certainly, it is not always easy to determine whether the revision of evaluations in some practice is tracking relevant environmental features – or rather aims just at aligning the sensibilities and responses of the participants of the practice. So, there are diverging views, for instance, about whether moral or aesthetic discourses are objective – about whether they aim at representing perspective-independent facts (see, for example, Bender 2003; Zangwill 2003; Schafer 2011; Stevenson 1963; Boyd 1988; Shafer-Landau 2003; Kalderon 2005b). As I mentioned in chapter 2, I will not engage in the debate about the objective character of these areas of discourse. My (modest, no doubt) claim is that our empirical views about everyday ‘middle-sized dry goods’ around us answer to the (perspective-independent) way the world is: in particular, our evaluations of the outcomes of our practical interactions with our surroundings tend to reflect relevant environmental features.

Why is this claim less controversial than the analogous proposal about aesthetic, moral or gastronomic discourse? On the one hand, in general there is far more consistency and agreement among different evaluative perspectives in the case of empirical views about the things around us than in the case of moral and aesthetic discourse – where radical divergences of judgment are more extended. True, our views about the empirical world have changed greatly (people once thought that the Earth was flat), but individuals typically agree on their assessment of the outcomes of simple practical interactions with their surroundings – for instance, in general it is easy to reach an agreement about whether James succeeded in turning the television on: we just have to see whether the television is on after James pressed the remote. This pervasive agreement among alternative perspectives would be explained if the assessing attitudes of individuals tended to track commonly detectable environmental trends and patterns.¹³² On the other hand, as I have argued above, there are plausible biological

¹³² MacFarlane (2007) argues that, in light of the pervasiveness of persistent disagreement in discourse about taste or humor, an objectivist position about those areas of discourse would force us to endorse a widespread skepticism about people’s ability to detect the relevant properties:

explanations of how our cognitive capacities have evolved to detect useful features of our physical surroundings.

To sum up: in the previous sections I claimed that the objectivity of representational practices requires that there is a distinction between being actually correct and merely being taken to be correct – and, in particular, a distinction between an action being actually successful and it merely being assessed as successful by some agent. When agents are fallible, this distinction requires that the evaluating attitudes of every agent or group of agents are taken to be revisable (so that error is always an open possibility). This, however, is only a necessary condition for objective representation, not a sufficient one. In order for objective representational practices to arise it is not only necessary that our attitudes are revisable, but also that their revision is to some extent sensitive to the way the world is. In the sort of practice I have been described, the process of revision of the attitudes of agents is driven by their evaluations of the outcomes of their practical interactions of the world. If the practice is to count as engaging in objective representation of the world, therefore, these evaluations may be, by and large, responsive to the way the world is. In this section I have argued that this minimal sensitivity is presupposed by human representational practices: if we are to see our practices as (objectively) representational, we have to assume that our evaluations track relevant environmental features. There is no reason, however, to think otherwise. Given the (plausible) biological origins of our practices and the fact that we seem to be well adapted to our surroundings, the most reasonable assumption is that our evaluations of practical success tend to be sensitive to the way the world is. If this is the case, then the resulting practices deserve to be counted as representational and subject to objective

If there are wholly objective properties of funniness, deliciousness, or likelihood, then most of us must be defective in our capacity to detect them. We must be humorblind, or taste-blind, or likelihood-blind, in much the same way that some of us are color-blind. But this diagnosis clashes with the way we think and talk about these domains. In our judgements about what is delicious, we lack the humility color-blind people show in their judgements about what is red or green. We do not seem to regard the fact that many others disagree with us as grounds for caution in calling foods delicious. We readily judge things to be funny in light of our own senses of humor, even though if challenged we can offer no grounds for thinking our senses of humor are the right ones.'

(MacFarlane 2007: 17; see Schafer 2011 for discussion).

standards of correctness. Our thought and talk can be said to be about the world, provided that we are minimally sensitive to the actual success of our actions.

8.6 Conclusion

The sort of pragmatist approach I am considering characterizes the representational purport of beliefs and assertions in terms of the success conditions of the actions licensed by them. Clearly, this approach can only provide a complete account of representation if such success conditions are shown to depend on how things are in the world. This is the question that has been discussed in this chapter.

According to the proposal I have put forward, the conditions of success of actions are determined by the evaluations carried out by agents. This proposal may, on the face of it, seem unable to account for the objectivity of our language and thought: ultimately, correctness would depend on our evaluative attitudes, and not on the way the world is.

I have tried to make room for objective representation in two ways. First, the notion of incorrectness may be articulated in practices with a multiplicity of evaluative perspectives (in line with Brandom's proposal (1994: ch.8)). Such multiplicity, I have argued, finds its most developed manifestation in social contexts.

Second, I have tried to argue that friction with the world is introduced if the evaluative responses of agents do track relevant environmental patterns. In organisms adapted to their environment, this will tend to be the case. Evolution has seen to it.

The resulting picture, I think, manages to characterize the kind of objectivity at play in human representational practices. Human discursive practices can be said to incorporate a representational dimension because agents update and revise their views in accordance with their evaluation of the success of their actions – and these evaluative responses typically depend on the way things are in the world, at least in agents minimally attuned to their environment. Moreover, any assessing attitude may be challenged by incompatible evaluations – including those made by other agents. As a result, assessing attitudes are defeasible: any attribution of correctness to a move of the practice may be retracted in light of suitable challenges. It is in virtue of this open-ended

Chapter 8

corrigible nature of human representational practices that objectivity – at least some form of objectivity – may find its place in them. Agents are always open, at least in principle, to the possibility of finding out that their views about the world are actually wrong.

Chapter 9: Conclusions

The main thesis of this dissertation is that discursive practices incorporate a representational dimension by virtue of involving interactions with the environment. However, practical engagement with the world is not a sufficient condition for representation: many non-representational practices involve dealings with the world (for instance, the practice of eating with knives and forks). I have examined, thus, what sort of interactions with the world allow for the emergence of objective representation – and I have discussed the main features of the resulting practices.

I have assumed as essential to the notion of representation that there is a distinction between correct (accurate) and incorrect (inaccurate) representations: correct representations accord with the way the things represented actually are, incorrect representations do not. In particular, objective representation requires that representational correctness depends on how the world represented is, rather than on the attitudes of those representing it. There must be a distinction between being actually correct and merely being regarded as correct from the perspective of any of the agents taking part in the representational practice.

I have also assumed that standards of correctness are established by the attitudes of agents. The challenge, thus, is to show how such attitudes can institute standards that override them – that is, standards according to which what is correct is determined by the way the world is, rather than by the attitudes of agents.

My strategy for facing this challenge has been to argue that although it is ultimately the attitudes of agents that establish standards of correctness, agents may defer to the world in order to determine what counts as correct. The world gets to have a say in what is correct because the participants in representational practices delegate such authority. More specifically, agents defer to the outcomes of their practical interactions with the world in order to assess the correctness of their representational views. Since the way the world is determines the outcomes of these interactions, standards of correctness in representational practices are constrained by the world. In this sense, it can be said that human representational practices answer to the way the world is.

Therefore, by virtue of the role that practical engagement with the environment play in the revision and updating of our representational views, discursive practices become responsive to the way the world is. However, the connection between the outcomes of practical interactions with the world and the conditions of correctness of specific beliefs is not straightforward: conditions of correctness cannot be directly derived from conditions of success. The reason for this is that unconsidered facts can always meddle with the outcomes of our interactions with the world; so, an action undertaken in light (only) of true beliefs may fail due to presence of unexpected obstacles. I have claimed that the possibility of such unconsidered impediments is reflected by the defeasibility of our representational practices, and in particular by the non-monotonicity of practical reasoning. Typically, agents are not in a position to rule out explicitly each possible defeater; thus, when actions are recommended as good means for achieving their goal, it is done under the assumption that (indefinitely many) defeating possibilities are not taking place. There is no need, however, to see these pieces of practical reasoning as enthymematic or somehow defective, rather than as (defeasibly) good. If the agent possesses no reasons to suspect that certain possible defeaters are taking place, her practical reasoning may remain good even if such defeating possibilities are not explicitly ruled out.

As a consequence of this defeasibility, the connection between the evaluation and practical outcomes and the revision of representational views is not piecemeal, but holistic. When an action fails, it can be because the beliefs recommending it were inaccurate or also because some unexpected obstacle was taking place – which, in the picture I have presented, would defeat the relevant piece of practical reasoning. Nevertheless, representational views are still sensitive to the outcomes of practical interactions with the world, even if it is whole webs of intertwined discursive commitments that have to be revised in light of such outcomes. Representational views are updated in systematic ways in response to the outcomes of open-ended practical dealings with the world.

I have argued that this defeasibility introduces an occasion-sensitive dimension in practices of giving and asking for reasons and, accordingly, in inferentialist semantics. This is to say, what counts as good reasons for what depends on features of the circumstances of the occasion of reasoning. In particular, such occasion-dependent factors determine

whether a given defeating possibility has to be ruled out explicitly, or can remain unconsidered, without making the agent's reasoning enthymematic (of course, that the possibility of a certain defeater may remain unconsidered does not mean that, if such a defeater turned out to obtain after all, the piece of reasoning would not be actually defeated). I have argued that an agent needs to consider the possibility of a given defeater only when there are reasons to do so – only when there are reasons to *suspect* that the defeater might obtain. Whether there are such reasons, I have claimed, is determined by occasion-sensitive factors, both pertaining to the situation the agent is reasoning about (e.g. what environmental conditions are typical in that situation) and to the context in which reasons are been given and asked for (e.g. what information is available in the context or how high stakes are). These considerations, I think, offer an illuminating angle on the occasion-sensitivity of reason-giving practices, and also reveal a largely unexplored form of occasion-sensitivity in inferentialist semantics. The study of this occasion-sensitivity is an interesting line of further research arising from the issues discussed in this dissertation.

Defeasibility is not only a feature of the connection between representational correctness and practical success, but also permeates agents' evaluations of practical success. Representational views are revised in light of the outcomes of actions *as assessed from the perspective of agents*. This assessment is fallible – agents could be mistaken in their evaluations of the outcomes of their actions. Acknowledging this fallibility does not imply, however, that agents cannot take their evaluations of practical outcomes to be determined by the way things actually are. Typically, agents will assume that, by and large, their evaluations succeed in tracking relevant environmental features – that is, that their evaluations respond to the way the world is. Unless they have special reasons for doing so, agents are not required to consider every possible defeater for their evaluations. When things go well and such possible defeaters are actually absent, there is no reason not to think of the agent's evaluations as reflecting the way the world is.

Defeasibility, therefore, is an essential aspect of human representational practices – at least of representational practices concerned with everyday empirical matters. It pervades such practices through and through: in particular, both the acquisition of perceptual entitlement (i.e. entitlements earned by virtue of the exercise of reliable perceptual capacities) and of practical entitlement (i.e. entitlement to performing a

certain action as a good means for achieving a goal) are typically defeasible. It must be noted, however, that defeasibility does not turn objective representation into an ideal unattainable in real practices. Rather, defeasibility is the form adopted by practices which are both subject to objective standards of correctness and established by fallible agents. By putting forward their views in a defeasible way, agents leave room for the open-ended possibility of such views being revealed as mistaken – as not reflecting how things actually are.

What I have described is the structure of practices that are taken to be objective by their participants. From the point of view of the participants, a practice that is actually subject to objective standards may not be distinguishable from a practice that only seems to be so (e.g. a practice that is actually relativist). In order for the practice to actually answer to the way things are – in order for the practice to be bound by objective standards –, it has to be suitably embedded in the world (in particular, it has to be embedded in the world in such a way that the evaluations of agents tend to track relevant environmental features). When this sort of practices is actually suitably intertwined with the environment, it can be said that it answers to the way the world is. The fact that agents cannot always tell the difference between situations in which their practices are actually responsive to the world and situation in which they only appear to be so does not mean that such practices should not count as responsive to the world in cases when they actually are.

My discussion has focused on our practices of representation concerning ordinary, empirical matters (i.e. matters such as the number of onions in my kitchen or the weather outside). I have only tangentially dealt with representation in other domains – for instance, representation in moral, aesthetic or mathematical discourse. One interesting question is whether objective representation is possible in practices that do not involve the sort of practical interactions with the world I have described; if the answer is affirmative, a further question is how such practices get to be constrained by the things represented. I have remained neutral about these questions. It may be that, absent the sort of practical interactions I have discussed, a practice cannot involve genuine objective representation – for instance, it could be that in such practices only non-representational attitudes are really expressed. An alternative possibility is that there is some story, differing from the one I have told, about how objective standards arise in these domains –

a story about how correctness in those practices gets to be sufficiently decoupled from the attitudes of agents. What it takes for agents to count as engaging in representational practices may be different in different domains of discourse. Exploring these questions is an interesting area of further research.

My aim, thus, has not been to offer a general account of representation, but only to explore to what extent our ordinary discourse about the empirical world can be said to be representational – more specifically, to what extent it can be said to be bound by objective standards of correctness. My account of what it is to engage in representational practices has not been constructed in terms of any specific word-world relation. However, nothing in the analysis I have presented prevents one from using semantic representational notions, or even from saying that claims stand in representational relations with the world they represent (for instance, one can say that some expression is used to refer to a certain object in the world). If my proposal counts as non-representationalist, it is because I have required that explanations of intentionality – explanations of what it is to engage in representational practices – should not be ultimately couched in terms of such word-world relations. Instead, I have tried to account for intentionality in terms of use – that is, I have tried to explain what sort of practices gives rise to intentionality (characterizing these practices in non-intentional ways). Although of course the relevant practices will typically involve many different relations with the world, there is no reason to think that such practices can be characterized in a systematic, unified way in terms of any specific type of word-world relation (at least, a relation that is not specified in non-intentional terms).

This dissertation, therefore, has been animated by a pragmatist spirit. I have investigated what people do when they talk and think about the way the world is.

List of References

- Alvarez, M. 2010. *Kinds of reasons: an essay in the philosophy of action*. Oxford: Oxford University Press.
- Anscombe, G. E. M. 1957. *Intention*. Oxford: Blackwell.
- Audi, Robert. 1989. *Practical Reasoning*. London: Routledge.
- Ayer, A. J. 1936. *Language, Truth, and Logic*. London: Gollancz.
- Bach, K. 1999, The Semantics Pragmatics Distinction: What is and Why it Matters., in K. Turner (ed.), *The Semantics-Pragmatics Interface From Different Points of View*. Elsevier.
- Bach, K. 2002. Semantic, Pragmatic. In J. Keim Campbell, M. O'Rourke, and D. Shier, (eds.), *Meaning and Truth*. New York: Seven Bridges Press.
- Barandarian, X., Egbert, N. 2014. Norm-establishing and norm-following in autonomous agency. *Artificial Life* 1, 1-24.
- Beisecker, D. 1999. The importance of being erroneous: prospects for animal intentionality. *Philosophical Topics*, 27, 281-308.
- Bender, J. 2003. Aesthetic realism 2. In Levinson, J. (Ed.), *The Oxford Handbook of Aesthetics*. Oxford: Oxford University Press.
- Bigelow, J., & Pargetter, R. 1987. Functions. *The Journal of Philosophy*, 84, 181-196.
- Blackburn, S. 1993, *Essays in Quasi-Realism*, New York: Oxford University Press.
- Blackburn, S., 1998, *Ruling Passions*, Oxford: Clarendon Press.
- Blackburn, S. 2010. *Practical tortoise raising: and other philosophical essays*. Oxford: Oxford University Press.
- Blackburn, S. 2013, Pragmatism: All or Some? In H. Price, *Expressivism, Pragmatism and Representationalism*, Cambridge: Cambridge University Press.
- Block, N. 1986: Advertisement for a semantics for psychology. *Midwest Studies in Philosophy* 10, 615-78.

Bibliography

- Blome-Tillmann, M. 2009. Knowledge and presuppositions. *Mind*, 118, 241-294.
- Boghossian, P. 2008: *Content and Justification*. Oxford: Oxford University Press.
- Boorse, C. 2002. A rebuttal on functions. In A. R. Ariew, R. Cummins and M. Perlman (eds.), *Functions*. Oxford: Oxford University Press.
- Boyd, R., 1988. How to be a Moral Realist. In G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca: Cornell University Press.
- Boyd, R., 2003a, Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part I. *Philosophy and Phenomenological Research*, 66, 505–553.
- Boyd, R., 2003b, Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part II. *Philosophy and Phenomenological Research*, 67, 24–47.
- Braddon-Mitchell, D., & Jackson, F. 1997. The teleological theory of content. *Australasian Journal of Philosophy*, 75, 474-489.
- Brandom, R. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.
- Brandom, R. 2000. *Articulating Reasons*. Cambridge, MA: Harvard University Press.
- Brandom, R. 1994b. Unsuccessful Semantics. *Analysis* 54, 175-178
- Brandom, R. 2000. *Articulating Reasons*. Cambridge, MA: Harvard University Press.
- Brandom, R. 2002. Pragmatics and Pragmatisms, in *Hilary Putnam: Pragmatism and Realism*, James Conant and Urszula M. Zeglen (eds.), London: Routledge.
- Brandom, R. 2008. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.
- Brandom, R. 2011. *Perspectives on Pragmatism*. Cambridge, MA: Harvard University Press.
- Broome, J. 1999. Normative Requirements. *Ratio* 12, 398-419.

- Brown, J. 2008. Subject-Sensitive Invariantism and the Knowledge Norm for Practical Reasoning. *Noûs*, 42, 167-189.
- Burge, T. 2009. Primitive Agency and Natural Norms. *Philosophy and Phenomenological Research*, 79, 251-278.
- Canfield, J. 1964. Teleological explanation in biology. *British Journal for the Philosophy of Science*, 14, 285-295.
- Chrisman, M. 2007. From epistemic contextualism to epistemic expressivism. *Philosophical Studies*, 135, 225-254.
- Chrisman, M. 2008. Expressivism, inferentialism, and saving the debate. *Philosophy and Phenomenological Research*, 77, 334-358.
- Chrisman, M. 2010a, Expressivism, Inferentialism and the Theory of Meaning. In MS Brady (ed.), *New waves in Metaethics*. Basingstoke: Palgrave Macmillan.
- Chrisman, M 2010b. From Epistemic expressivism to epistemic inferentialism. In A Haddock, A Millar & D. Pritchard (eds), *Social Epistemology*. Oxford: Oxford University Press.
- Chrisman, M. (2012). On the Meaning of 'Ought'. *Oxford Studies in Metaethics*, 7, 304-332.
- Christensen, W.d., and Bickhard, M. H. 2002. The Process Dynamics of Normative Function. *The Monist*, 85, 3-28.
- Dancy, J. 2000. *Practical Rationality*. Oxford: Oxford University Press.
- Dancy, J. 2004. *Ethics without Principles*. Oxford: Oxford University Press.
- Del Corral, M. 2013. Against Normative Judgement Internalism. *Organon F*, 20, 567-587.
- Delancey, C. 2006. Ontology and Teleofunctions: A Defense and Revision of the Systematic Account of Teleological Explanation. *Synthese*, 150, 69–98.
- DeRose, K. 1992. Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, 52, 913-929.
- Dewey, J. 1896. The reflex arc concept in psychology. *Psychological review*, 3, 357-370.

Bibliography

- Di Paolo, E. 2005: Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4, 429-452.
- Dokic, J. and Engel, P. 2002. *Frank Ramsey: Truth and Success*. London: Routledge.
- Dokic, J. and Engel, P. 2005. Ramsey's Principle Re-situated. In Lillehammer, H. And Mellor, H. (eds.) *Ramsey's legacy*, Oxford: Oxford University Press.
- Dreier, J. 2004. Meta-Ethics and the Problem of Creeping Minimalism. *Philosophical Perspectives*, 18, 23-44.
- Dretske, F. 1995 *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Dretske, F., 1981, *Knowledge and the Flow of Information* , Cambridge, MA: MIT Press.
- Dummett, M. 1981. *Frege: Philosophy of Language* (2nd edn). Cambridge, MA: Cambridge University Press.
- Dummett, M. A. 1993. *The seas of language*. Oxford: Clarendon Press.
- Fantl, J., & McGrath, M. 2002. Evidence, pragmatics, and justification. *Philosophical Review*, 67-94.
- Fantl, J., & McGrath, M. 2007. On pragmatic encroachment in epistemology. *Philosophy and Phenomenological Research*, 75, 558-589.
- Field, H., 1980, *Science Without Numbers*, Princeton: Princeton University Press.
- Field, H., 1989, *Realism, Mathematics and Modality*, Oxford: Blackwell.
- Fodor, J. 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press, Bradford Books.
- Fodor, J. 1990, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, Bradford Book.
- Fodor, J. and Lepore, E. 2001. Brandom's Burdens. *Philosophy and Phenomenological Research*, 63 , 465-481.

- Fodor, J. and Lepore, E. 2007. Brandom Beleaguered. *Philosophy and Phenomenological Research*, 74 , 677-691.
- Gangopadhyay, N., & Kiverstein, J. 2009. Enactivism and the unity of perception and action. *Topoi*, 28, 63-73.
- Gauker, C. 2003 *Words without meaning*. Cambridge, MA: MIT Press.
- Gibbard, A., 1990, *Wise Choices, Apt Feelings*, Cambridge: Harvard University Press.
- Gibbard, A., 2003, *Thinking How to Live*, Cambridge: Harvard University Press.
- Gibbons, J. 2010. Things that make things reasonable. *Philosophy and Phenomenological Research*, 81, 335-361.
- Glüer, K. and Å. Wikforss, 2009, Against Content Normativity, *Mind*, 118, 31–70.
- Grönert, P., 2005. Brandom's Solution to the Objectivity Problem, in: *Pragmatics and Cognition*, 13, 161-175.
- Hare, R.M. 1952. *The Language of Morals*. Oxford: Oxford University Press.
- Harman, G. 1976. Practical reasoning. *The Review of Metaphysics*, 431-463.
- Harman, G. 1982: Conceptual Role Semantics. *Notre Dame Journal of Formal Logic*. 23, 242-256.
- Harman, G. 1984. Logic and reasoning. *Synthese*, 60, 107-127.
- Harman, G. 1986. *Change in View*. Cambridge, MA: MIT Press.
- Hattiangadi, A. 2006. Is Meaning Normative? *Mind and Language*, 21, 220–240.
- Hattiangadi, A. 2007. *Oughts and Thoughts: Scepticism and the Normativity of Content*. Oxford: Oxford University Press.
- Hattiangadi, A. 2009. Some more thoughts on semantic oughts: a reply to Daniel Whiting. *Analysis*, 69, 54-63.
- Haugeland, J. 1982. Heidegger on being a person. *Nous*, 16, 15-26.

Bibliography

- Haugeland, J. 1998. *Having thought. Essays in the Metaphysics of Mind*. Cambridge, MA: Cambridge University Press.
- Horwich, P. 1990. *Truth*. Oxford: Oxford University Press.
- Horwich, P. 1998. *Meaning*. Oxford: Oxford University Press.
- Horwich, P. 2005. *Reflections on meaning*. Oxford: Oxford University Press.
- Horwich, P. 2008. A New Framework For Semantics. *Philosophical Perspectives*, 22, 233-240.
- Hurley, S. 2001. Perception and action: Alternative views. *Synthese*, 129, 3-40.
- Hussain, N.J.Z. 2004. The return of moral fictionalism. *Philosophical Perspectives* 18, 149-187.
- Jackson, F. 1998, *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Clarendon Press.
- Jonas, H. 1968. Biological foundations of individuality. *International Philosophical Quarterly*, 8, 231–251.
- Joyce, R., 2001, *The Myth of Morality*, Cambridge: Cambridge University Press.
- Kalderon, M., (ed.), 2005a, *Fictionalism in Metaphysics*, Oxford: Clarendon Press.
- Kalderon, M., 2005b, *Moral Fictionalism*, Oxford: Clarendon Press.
- Kölbel, M. 2015a. Relativism 1: Representational Content. *Philosophy Compass*, 10, 38-51.
- Kölbel, M. 2015b. Relativism 2: Semantic Content. *Philosophy Compass*, 10, 52-67.
- Kukla, R., & Lance, M. 2009. *'Yo!'and'Lo!': The Pragmatic Topography of the Space of Reasons*. Cambridge, MA: Harvard University Press.
- Kukla, R., & Lance, M. 2014. Intersubjectivity and Receptive Experience. *The Southern Journal of Philosophy*, 52, 22-42.
- Laurier, D., 2008: Pragmatics, Pittsburgh Style. In Stekeler-Weithofer, Pirmin (ed.), *The Pragmatics of Making it Explicit: On Robert B. Brandom*. Amsterdam: John Benjamins B. V.

- Laurier, D. 2009. Making 'Reasons' Explicit: How Normative is Brandom's Inferentialism? *Abstracta* 5, 79-100
- Levine, S. 2008. Rorty, Davidson, and the New Pragmatists. *Philosophical Topics*, 36, 167-192.
- Levine, S. 2010. Rehabilitating objectivity: Rorty, Brandom, and the new pragmatism. *Canadian Journal of Philosophy*, 40, 567-589.
- Lewis, D. 1983. New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343-377.
- Lewis, D. 1996. Elusive knowledge. *Australasian journal of Philosophy*, 74, 549-567.
- Lewis, D. 1998. *Papers in Philosophical Logic*. Cambridge: Cambridge
- Littlejohn, C. 2012. *Justification and the truth-connection*. Cambridge: Cambridge University Press.
- Loewer, B. 1997. A guide to naturalizing semantics. In Wright, C. and Hale, B. (eds.) *A Companion to the Philosophy of Language*, Oxford: Blackwell, 108-126.
- López de Sa, D. L. 2010. The makings of truth: realism, response-dependence, and relativism. In C. Wright & N. Pedersen (eds.), *New Waves in Truth*, Basingstoke: Palgrave Macmillan, 191-204.
- MacFarlane, J. 2005. The assessment sensitivity of knowledge attributions. *Oxford studies in epistemology*, 1, 197-233.
- MacFarlane, J. 2007. Relativism and Disagreement. *Philosophical Studies* 132, 17-31.
- MacFarlane, J. 2010. Pragmatism and Inferentialism. In B. Weiss and J. Wanderer (eds.), *Reading Brandom*. London: Routledge.
- MacFarlane, J. 2014. *Assessment sensitivity: Relative truth and its applications*. Oxford: Oxford University Press.
- Mackie, J., 1977, *Ethics: Inventing Right and Wrong*, Harmondsworth, New York: Penguin.
- McDowell, J. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.

Bibliography

- McDowell, J. 1998 *Mind, Value and Reality*. Cambridge, MA: Harvard University Press
- McDowell, J. 2005. Motivating inferentialism. *Pragmatics & Cognition*, 13, 121-140.
- McKenna, R. 2014. Normative scorekeeping. *Synthese*, 191, 607-625
- Mellor, D. H. 1991. *Matters of metaphysics*. Cambridge: Cambridge University Press.
- Miller, GA, Galanter, E., & Pribram, KA (1960). *Plans and the structure of behavior*. New York: Holt, Rhinehart, & Winston.
- Millikan, R. G. 1984. *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT press.
- Millikan, R. G. 1989 Biosemantics, *The Journal of Philosophy*, 86, 281-297
- Millikan, R. G. 1990, Truth Rules, Hoverflies, and the Kripke-Wittgenstein paradox. *The Philosophical Review*, 99, 323-353
- Millikan, R. G. 2005. *Language: A biological model*. Oxford: Oxford University Press.
- Misak, C. J. (Ed.) 2007. *New pragmatists*. Oxford: Oxford University Press.
- Moore, G. E., 1903. *Principia Ethica*, Cambridge: Cambridge University Press.
- Mossio, M., Saborido, C., & Moreno, A. 2009. An organizational account of biological functions. *The British Journal for the Philosophy of Science*, 60, 813-841.
- Nagel, E. 1961. *The Structure of Science*. New York: Harcourt, Brace & World.
- Nanay, B. 2013. Success Semantics: The Sequel. *Philosophical Studies*, 165, 151-165
- Neander, K. 1991. Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science*, 58, 168-184.
- Neander, K. 1995, Malfunctioning and Misrepresenting, *Philosophical Studies*, 79, 109–141.
- Nietzsche, F. 1974. *The Gay Science*. W. Kaufmann (Trans.). New York: Vintage Books.
- Noë, A. 2004. *Action in perception*. Cambridge, MA: MIT press.

- Nolan, D., G. Restall and C. West, 2005. Moral Fictionalism Versus the Rest. *Australasian Journal of Philosophy*, 83, 307–30.
- O'Regan, J. K., & Noë, A. 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24, 939-973.
- Papineau, D. 1987, *Reality and Representation*, New York: Blackwell.
- Papineau, D. 1993, *Philosophical Naturalism* Oxford: Blackwell.
- Parfit, D. 2011. *On What Matters*. Oxford: Oxford University Press.
- Peirce, C. S. 1998 *The Essential Peirce, Selected Philosophical Writings, Volume 2 (1893–1913)*, Bloomington, IN: Indiana University Press.
- Pinker, S. 1997. *How the Mind Works*, New York. Norton.
- Price, H. 2011a, *Naturalism Without Mirrors*, Oxford: Oxford University Press.
- Price, H. 2011b. Expressivism for two voices. In J. Knowles & H. Rydenfelt, eds., *Pragmatism, Science and Naturalism*. Zürich: Peter Lang, Zürich, 87—113.
- Price, H. 2013, *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.
- Price, H. 2014. From quasirealism to global expressivism — and back again? Forthcoming in R. Johnson and M. Smith (eds.), *Passions and Projections: Themes from the Philosophy of Simon Blackburn*. Oxford: Oxford University Press.
- Prien, B. 2010. Robert Brandom on Communication, Reference, and Objectivity. *International Journal of Philosophical Studies*, 18, 433-458.
- Ramberg, B. 2000. Post-Ontological Philosophy of Mind: Rorty versus Davidson. In Robert Brandom (Ed.), *Rorty and His Critics*. Oxford: Blackwell.
- Ramsey, F. 1927, Facts and Propositions. *Proceedings of the Aristotelian Society*, Supplementary Volumes, 7, 153-206
- Reed, B. 2010. A defense of stable invariantism. *Noûs*, 44, 224-244.

Bibliography

- Rödl, S., 2010, Normativity of Mind versus Philosophy as Explanation., in in B. Weiss and J. Wanderer (eds), *Reading Brandom*, 2010, New York: Routledge.
- Rorty, R. 1979. Philosophy and the mirror of nature. *Princeton, NJ: Princeton University Press*.
- Rorty, R. 1982. *Consequences of pragmatism: Essays, 1972-1980*. U of Minnesota Press.
- Rorty, R. 1991. Objectivism, Relativity, and Truth. *Philosophical Papers, 1*.
- Rorty, R. 2003. Comments on Jeffrey Stout's 'Democracy and Tradition'. Paper delivered at the Annual Meeting of the American Academy of Religion, Atlanta, 23 November 2003.
- Rosen, G. 1997. Who makes the rules around here?. *Philosophy and Phenomenological Research, 57*, 163-171.
- Rosen, G. 2001. Brandom on modality, normativity and intentionality. *Philosophy and Phenomenological Research, 63*, 611-623.
- Rosenblueth, A., Wiener, N., Bigelow, J. 1943. Behavior, Purpose and Teleology. *Philosophy of Science, 10*, 18-24.
- Ruse, M. E. 1971. Functional statements in biology. *Philosophy of Science, 38*, 87-95.
- Saborido, C., Mossio, M. and Moreno, A. 2011. Biological Organization and Cross-Generation Functions. *British Journal for the Philosophy of Science, 62*, 583-606.
- Schafer, K. 2011. Faultless disagreement and aesthetic realism. *Philosophy and Phenomenological research, 82*, 265-286.
- Schaffner, K. 1993. *Discovery and Explanation in Biology and Medicine*. Chicago: The University of Chicago Press.
- Scheer, J. K. (Ed.). 2013. *Mind, Reason and Being-in-the-World: the McDowell-Dreyfus debate*. New York: Routledge.
- Schnieder, B. 2010. Expressivism concerning epistemic modals. *The Philosophical Quarterly, 60*, 601-615.

- Schroeder, M. 2008. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
- Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Sellars, W. 1953. Inference and Meaning. *Mind*, 62, 313-338.
- Sellars, W. 1954. Some reflections on language games. *Philosophy of Science*, 204-228.
- Sellars, W. 1956. Empiricism and the Philosophy of Mind. *Minnesota studies in the philosophy of science*, 1, 253-329.
- Shafer-Landau, R. 2003. *Moral Realism: A Defence*, Oxford: Oxford University Press.
- Shapiro, L. 2004. Brandom on the Normativity of Meaning. *Philosophy and Phenomenological Research*, 68, 141-160.
- Sober, E. 1993. Mathematics and Indispensability, *Philosophical Review*, 102, 35–57.
- Stalnaker, R. 2002. Common ground. *Linguistics and philosophy*, 25, 701-721.
- Stanley, J. 2005. *Knowledge and practical interests*. Oxford: Oxford University Press.
- Stevenson, C.L. 1944. *Ethics and Language*. New Haven CT: Yale University Press.
- Stevenson, C.L. 1963. *Facts and Values*. New Haven CT: Yale University Press.
- Stout, J. 2007. On our interest in getting things right: Pragmatism without narcissism. In Ch. Misak (Ed.) *New Pragmatists*. Oxford: OUP, 7-31.
- Streumer, B. 2010. Practical reasoning. In Timothy O'Connor and Constantine Sandis (eds.) *A Companion to the Philosophy of Action*, Oxford: Wiley-Blackwell, 244-251.
- Varela, F. J. 1997. Patterns of life: Intertwining identity and cognition. *Brain and cognition*, 34, 72-87.
- Vogler, C. 2001. Anscombe on practical inference. In Elijah Millgram (Ed.) *Varieties of Practical Reasoning*, Cambridge, MA: MIT Press, 437-464.
- Wanderer, J. 2008. *Robert Brandom*. Montreal: McGill Queens University Press.

Bibliography

- Weber, A., Varela, J.. 2002. Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences* 1, 97-125.
- White, H. 2003. Brandom on Practical Reason. *Noûs*, 32, 127-139.
- Whiting, D., 2007. The normativity of meaning defended, *Analysis*, 67, 133–40.
- Whiting, D. 2009. Is meaning fraught with ought?. *Pacific Philosophical Quarterly*, 90, 535-555.
- Whiting, D. 2010. Should I believe the truth?. *Dialectica*, 64, 213-224.
- Whiting, D. 2013. Stick to the facts: On the norms of assertion. *Erkenntnis*, 78, 847-867.
- Whyte, J. T. 1990. Success Semantics. *Analysis* 50, 149-157
- Whyte, J. T. 1991. The Normal Rewards of Success. *Analysis* 51 2
- Whyte, J. T. 1993, Purpose and Content, *The British Journal for the Philosophy of Science*, 44, 45-60
- Whyte, J. T. 1997. Success Again: replies to Brandom and Godfrey-Smith. *Analysis*, 57, 84-88.
- Williams, B. 1985. *Ethics and the limits of philosophy*. London: Fontana.
- Williams, M. 1991. *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*. Oxford: Basil Blackwell.
- Williams, M. 2001. *Problems of knowledge: A critical introduction to epistemology*. Oxford: Oxford University Press.
- Williams, M. 2013, How pragmatists can be local expressivists. In H. Price, *Expressivism, Pragmatism and Representationalism*, Cambridge: Cambridge University Press, 128-144
- Williamson, T., 2000. *Knowledge and its Limits*, Oxford: Oxford University Press.

Williamson, T., 2003. Understanding and Inference. *Aristotelian Society Supplementary Volume 77* (1):249–293. Williamson, T. 2009. Reference, inference and the semantics of pejoratives. *The Philosophy of David Kaplan*, 137-158.

Wittgenstein, L. 1979 *Notebooks 1914-1916*, eds. G.H. von Wright & G.E.M Anscombe; tr. G.E.M. Anscombe. Oxford: Blackwell.

Wittgenstein, L., 1953. *Philosophical Investigations (PI)*. Oxford: Blackwell.

Wright, C. 2001. *Rails to infinity: essays on themes from Wittgenstein's Philosophical investigations*. Cambridge, MA: Harvard University Press.

Yalcin, S. 2011. Nonfactualism about epistemic modality. In A. Egan.& B. Weatherson (Eds.). *Epistemic modality*. Oxford: Oxford University Press.

Yalcin, S. 2012. Bayesian Expressivism. *Proceedings of the Aristotelian Society 112*, 123-160.

Zangwill, N. 2003. Aesthetic realism I. In Levinson, J. (Ed.), *The Oxford Handbook of Aesthetics*. Oxford: Oxford University Press, 63-79.

Zangwill, N. 2003. Aesthetic realism I. In Levinson, J. (Ed.), *The Oxford Handbook of Aesthetics*. Oxford: Oxford University Press, 63-79.