

# UKUTOOMING

The Communicative Significance  
of Beliefs and Desires



DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS

II

DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS

II

**UKU TOOMING**

The Communicative Significance  
of Beliefs and Desires



UNIVERSITY OF TARTU  
PRESS

Dissertation has been accepted for defence of the degree of Doctor of Philosophy (PhD) in Philosophy in August 18, 2014 by the Council of the Institute of Philosophy and Semiotics, University of Tartu

Supervisors: Dr. Bruno Mölder

Opponents: Prof. Dr. Christopher Gauker  
(University of Salzburg, Austria)

Dr. Julia Tanney  
(University of Kent, United Kingdom)

Defence: the thesis will be defended at the University of Tartu, Estonia,  
on September 24, 2014, at 13.00, in the Senate room

This thesis has been published with the support of European Union through the European Social Fund (Graduate School of Linguistics, Philosophy and Semiotics)



European Union  
European Social Fund



Investing in your future

ISSN 1406-9520  
ISBN 978-9949-32-656-3 (print)  
ISBN 978-9949-32-657-0 (pdf)

Copyright: Uku Tooming, 2014

University of Tartu Press  
[www.tyk.ee](http://www.tyk.ee)

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	8
1. INTRODUCTION .....	9
1.1 Beliefs and desires .....	9
1.2 The question .....	10
1.3 Outline of the thesis .....	13
1.4 Concluding remarks .....	14
2. THE IMPORTANCE OF AFFECTABILITY .....	16
2.1 Introduction .....	16
2.2 The standard picture .....	16
2.3 On the predictive role of folk psychology .....	19
2.4 Alternative: mindshaping .....	23
2.5 Evaluating the mindshaping hypothesis .....	28
2.6 Affectability – engaged mindreading .....	29
2.7 The centrality of engaged responses .....	32
2.7.1 Independence from predictions and mindshaping .....	32
2.7.2 Distinguishing the role of affectability .....	33
2.7.3 The indispensability of engaged responses .....	34
2.8 Conclusion .....	36
3. MORE ABOUT AFFECTABILITY: FROM ADOPTION TO NEGOTIATION .....	38
3.1. Introduction .....	38
3.2 Mimetic belief, mimetic desire .....	39
3.2.1 The general idea .....	39
3.2.2 Mimetic belief .....	41
3.2.3 Mimetic desire .....	44
3.3 Reactive attitudes and folk psychology .....	47
3.4 Manipulating and negotiating attitudes .....	53
3.4.1 Becoming active .....	53
3.4.2 Manipulation of attitudes .....	55
3.4.3 Negotiating attitudes .....	57
3.5 Taking stock .....	60
4. FROM AFFECTABILITY TO THE COMMUNICATIVE CONCEPTION .....	63
4.1 Introduction .....	63
4.2 Initial proposals .....	65
4.3 From engaged responses to speech acts .....	69
4.4 The communicative conception .....	74
4.5 The communicative conception and the other benefits of folk psychology .....	77
4.6 On behalf of the communicative conception .....	81
4.7 Conclusion .....	84

5. MODIFICATION OF THE COMMUNICATIVE CONCEPTION .....	85
5.1 Introduction .....	85
5.2 Problems with the communicative conception .....	85
5.2.1 The initial worry .....	85
5.2.2 Conceptual need for mentalistic concepts .....	87
5.2.3 The solution? .....	92
5.3 The modified communicative conception .....	93
5.3.1 Desires .....	93
5.3.2 Beliefs .....	97
5.4 Comparisons .....	103
5.4.1 Desires and the hedonic theory .....	103
5.4.2 Beliefs and deontic scorekeeping .....	104
5.5 Problems .....	107
5.5.1 Language first? .....	107
5.5.2 Right kind of attitudes? .....	108
5.5.3 Different kinds of desire .....	108
5.5.4 Different kinds of belief .....	111
5.6 Conclusion .....	113
6. THE COMMUNICATIVE CONCEPTION AND ONTOLOGY .....	115
6.1 Introduction .....	115
6.2 From epistemology to ontology .....	116
6.3 Two ways to understand folk commitments .....	119
6.3.1 Lewis/Jackson view .....	119
6.3.2 The pragmatic view .....	121
6.4 Against the Lewis/Jackson view .....	124
6.5 Problems with the pragmatic view and the solution .....	127
6.6 The ontological status of beliefs and desires .....	130
6.6.1 Initial proposal: interpretivism .....	130
6.6.2 Interpretability and the modified communicative conception .....	136
6.6.3 The communicative conception and motivations for interpretivism .....	137
6.6.4 “Realistic Dennett” and the communicative conception .....	139
6.6.5 Comparison: Schwitzgebel .....	142
6.6.6 Affectability and ontology .....	145
6.6.7 Ontological quietism? .....	146
6.7 Conclusion .....	147
7. THE FIRST-PERSON PERSPECTIVE .....	149
7.1 Introduction .....	149
7.2 The phenomenology of beliefs and desires .....	150
7.2.1 Affectability and self-ascriptions .....	150
7.2.2 Alienated attitudes .....	154
7.2.3 Cognitive phenomenology? .....	157
7.2.4 Desires and the apparent good .....	161
7.3 The epistemology of self-ascriptions .....	163

7.3.1 From ascribing to knowing .....	163
7.4 Objections? .....	170
7.4.1 Two kinds of self-knowledge? .....	170
7.4.2 Desires and satisfaction .....	172
7.5 Comparison: The Interpretivist Sensory-Access account of self-knowledge .....	175
7.6 Conclusion .....	179
8. CONCLUSION .....	181
REFERENCES .....	187
SUMMARY IN ESTONIAN .....	199
CURRICULUM VITAE .....	204
ELULOOKIRJELDUS .....	206

## ACKNOWLEDGEMENTS

During the process of writing this thesis I have received support from many people. First of all, Bruno Mölder has been very supportive as my supervisor and has always encouraged me to pursue my projects, not only with regard to my thesis but also to other academic endeavours. I also thank other people at the Department of Philosophy in Tartu for providing a stimulating research environment. My gratitude goes to my fellow graduate students, especially to Riin Kõiv and Taavi Laanpere, with whom it was possible to discuss philosophy in a more informal setting. I also thank all of those who participated in the Analytic Philosophy Seminar and Summer Camp which provided an excellent opportunity to mingle and share research. In addition, I would like to express my gratitude to Jaan Kivistik who was my MA supervisor during my three-year foray into anthropology of religion and who helped to keep my interest in philosophy alive.

Much of the research that resulted in this thesis owes to my study abroad. During my stay in the University of Hertfordshire in autumn 2012 I profited from numerous discussions with Dan Hutto whose passion for philosophy was unmatched and whose family showed me great hospitality. I also got to know other philosophy students in Hertfordshire among whom I want to mention Zuzanna Rucinska and Dan Woolley who were great to converse with. My stay in Radboud University Nijmegen in spring 2013 benefited above all from conversations with Marc Slors whose comments on my work were extremely helpful and who involved me in the research activities there. I also thank the participants in the Mind & Action Seminar, especially Leon de Bruin and Fleur Jongepier, who provided valuable feedback.

I would also like to mention Søren Overgaard whose kind words about my poster in Bochum motivated me to write it into a full-length paper and Tad Zawidzki whose comments on the early version of the second chapter made me reconsider what I wanted to achieve there. I am indebted to Alex Davies who proofread the whole manuscript.

Finally, I give thanks to my family and friends without whom nothing of this could have been possible.

My research was supported by the Estonian Science Foundation Grant No. ETF9117.



# I. INTRODUCTION

## I.1 Beliefs and desires

This thesis is about intentional attitudes. These are mental states which are supposed to represent possible and actual states of affairs. In philosophy of mind, intentional attitudes (or propositional attitudes)<sup>1</sup> are usually divided into two types: cognitive and conative attitudes. Beliefs have been taken to be the paradigmatic cognitive attitudes and desires the paradigmatic conative ones. Both beliefs and desires involve two components – an attitude part and a content part – and have satisfaction conditions. Beliefs are satisfied when they are true and desires are satisfied when they are fulfilled (Fodor 1985: 78). The difference between these two attitudes may also be expressed in terms of “direction of fit”: while beliefs are supposed to conform to the world, desires aim at fitting the world to themselves (Platts 1979: 257). Spelled out in such general terms, this proposal remains rather vague. Another possibility would be to articulate the difference by arguing that only desires can directly motivate behaviour while beliefs only rationalize it (Smith 1987). This claim is rather controversial (Price 1989). A third option is to distinguish between these attitudes in terms of a relation they bear to their content, *p*. In the case of beliefs, the relation to content is holding-*p*-to-be-true. Desires, on the other hand, have the relation wanting-*p*-to-be-true to their content. Of course, this characterization is rather uninformative because the description of these relations already makes reference to beliefs and desires. After all, believing that *p* is the same as holding *p* to be true and desiring that *p* is the same as wanting *p* to be true. On a pretheoretical level it should still be clear that these terms capture certain distinctive areas of our linguistic practice. Because of that, speaking about others in terms of their beliefs and desires can be distinguished from other kinds of discourse. For instance, it can be distinguished from speaking about others’ social roles or about their digestive systems.

The most common questions that arise in philosophy of mind about intentional attitudes can be expressed as follows: what are beliefs and desires and what are the conditions for having these attitudes? The widespread assumption, which I also share, is that our conception of mental states should comport with our best psychological theories about the human mind. I am not going to analyse different theories of mind here but some distinctions can be made. For instance, some authors try to start from psychology and biology and then ask whether the best theories in those areas require postulating something similar to propositional attitudes (Sterelny 2003; Bermúdez 2003a). A more traditional approach begins with our folk conception of those states and then asks how beliefs and desires should be conceptualized and how they could be related to

---

<sup>1</sup> The idea that all intentional attitudes are exhausted by propositional attitudes is by no means uncontroversial (cf. Montague 2007). In the present thesis, I will nonetheless assume that beliefs and desires can be analysed in terms of relations to propositions.

scientifically respectable entities. The relevant relation can be identity or multiple realisability or something else (cf. Kim 2005). In any case, there is quite a lot of room for different theories when it comes to the relation between scientific and folk views. However, the folk conception of beliefs and desires plays an important part in all the approaches. Minimally, it can be taken to be the motivator for asking about the nature of mental states in the first place. Even if the question is simply how the mind works, philosophical approaches still ask whether our folk conception captures something similar to what scientific theories claim. Also, the folk conception usually offers an initial vocabulary in terms of which to even start asking what the mind is. After all, the notions of belief and desire originate from everyday talk.

## 1.2 The question

The fact that the folk conception is always the basis of our investigations of the mind, at least in some sense, motivates the central question of this thesis. The question is: what do we do with the concepts of belief and desire and what are the main benefits of their application? I am interested in what is distinctive about folk psychology<sup>2</sup> when we compare it to other parts of our conceptual machinery. The answer to this question also lets us see what it is about belief and desire attributions that makes these benefits possible.

How concepts of different mental states are acquired and which cognitive mechanisms enable their application is a widely researched topic in developmental psychology and philosophy of mind (Wellman 1990; Stich and Nichols 1993; Baron-Cohen 1995; Gopnik and Meltzoff 1997; Gordon 1986; Goldman 1989; Heal 2003; Stich and Nichols 2003; Currie and Ravenscroft 2002; Goldman 2006; Apperly 2011). As for what we do when we attribute beliefs and desires, the frequent, albeit often not explicit, assumption in the literature on folk psychology is that by attributing beliefs and desires to others we explain and predict their behaviour.<sup>3</sup> An alternative view which is defended by some authors is that instead of, or at least besides explaining and predicting, belief and desire attributions enable us to shape the minds and behaviours of others (McGeer 2007; Zawidzki 2008). I will argue that there is a third aspect to folk psychology which hasn't been considered before in the philosophical and psychological literature. There are certain ways in which attributers of belief and desire are affected by their attributions, most notably, by having the chance to agree or disagree with the attributed belief and to endorse or disapprove of the attributed desire. I will defend the claim that these ways of being affected (I label them with the term "affectability") are indispensable to folk psychology.

---

<sup>2</sup> What I mean by "folk psychology" is simply the way we think and speak about others in terms of beliefs and desires. I am here agnostic about whether folk psychology is a theory or not (cf. Stich & Ravenscroft 1996: 116).

<sup>3</sup> I will provide references in the next chapter.

Methodologically, the present approach, with its aim to clarify the use of certain concepts, has its origin in the philosophy of Gilbert Ryle (1949; 1962). It can also be related to the later Wittgenstein (1984).<sup>4</sup> Both philosophers based their work on observations of the ways in which concepts are applied in everyday practice and the concrete implications these applications have. This should allow us to step back from explanations that are given to folk psychological activities and to ask what those activities actually consist in. As is well known, both Ryle and Wittgenstein hoped that their endeavours would have a therapeutic effect on the way in which we see ourselves. To a modest extent, I want to achieve something similar with the present thesis. The fact that certain ways of being affected by mental state attributions have been overlooked in the study of folk psychology has resulted in a distorted picture of social cognition. In the present day, the clarificatory approach has been developed and defended with regard to folk psychological vocabulary by such authors as Julia Tanney, Daniel Hutto and Vincent Descombes. Descombes argues that mental states should be modelled on moves in rhetoric, not on physical causes as is usually done in philosophy of mind, because rhetoric as an art of persuasion reflects better what people are actually doing when they are attributing beliefs and desires (Descombes 2001: 90). Hutto observes that our understanding of other people usually depends on their concrete circumstances and relies on shared knowledge of how it is proper to act and respond. These observations are meant to give support to the idea that the conceptualisation of folk psychology as a detached and theoretical activity is on the wrong track (Hutto 2009: 206). Tanney criticises the cognitivist view of mind, according to which thinking consists of the manipulation of mental representations, by bringing out how this doesn't correspond to how people actually attribute mental states to one another. The attribution of mental states, according to Tanney, serves to articulate more precisely the circumstances of the attributee; there is no commitment to mental representations (Tanney 2005: 347). These authors each focus on the everyday employment of mental state concepts, on the basis of which they draw conclusions about the distinctive role of these concepts. This thesis uses similar observations to argue for the indispensability of affectability when people attribute beliefs and desires to one another. My account also draws inspiration from Huw Price who has suggested that mentalistic vocabulary may function in a way that distinguishes it from other areas of discourse by making a particular difference to the linguistic behaviour of those who use it (Price 1992; O'Leary-Hawthorne & Price 1996).<sup>5</sup> He himself has given an account of the concept of truth along these lines (Price 1988). However, he hasn't really

---

<sup>4</sup> The exegesis of Wittgenstein's philosophy is a tricky issue. Here I only want to point some similarities at a very general level.

<sup>5</sup> Price relates this suggestion historically to Ryle (Price 2007: 396) but also calls his proposal "broadly Kantian examination and classification of different domains of discourse, with reference to their role in human life" (Price 1992: 409). Since my thesis isn't historically oriented I won't speculate how far into history we can look to find the inklings of the present approach.

developed his suggestion with regard to mental state concepts much further. This thesis will do that.

I don't intend to confine myself to merely describing the central roles of folk psychological attribution within human social life, though. The analysis of the roles of belief and desire attribution can also indicate what these attributions must consist in for the respective concepts to fulfil these roles. This brings us to the second aim of my thesis. After having considered the main beneficial effects of the attribution of beliefs and desires, I will ask how we should understand these attributions in order to explain how these effects are possible. This means stating necessary conditions for someone to count as an attributer of beliefs and desires. Such conditions are meant to articulate what should be constitutive of folk psychology if it is to play the roles which we treat as valuable. My claim will be that in order to explain the indispensability of specific responses to attributed attitudes (agreement or disagreement in the case of beliefs and endorsement or disapproval in the case of desires), belief and desire attributions should involve reference to speech acts.

This way of inquiring into the conditions of mental state attribution is methodologically indebted to Justin Fisher's project of pragmatic conceptual analysis which also starts from considering the main beneficial effects that the use of a concept has. It then asks how application of the concept makes these effects possible. It should then be possible to define such conditions of concept-application which would deliver those beneficial effects in the most optimal way (Fisher 2006: 19). Even if proposed application conditions are unintuitive, pragmatic analysis can still be considered successful when the adoption of these conditions enables us to reap the benefits of the concept better than when the concept is employed with more intuitive application conditions (Ibid. 74).<sup>6</sup> There are other authors who have moved from an analysis of the benefits of a concept to a Fisher-style postulation of its application conditions. For instance, take Edward Craig who has proposed the following treatment of the concept of knowledge:

We take some *prima facie* plausible hypothesis about what the concept of knowledge does for us, what its role in our life might be, and then ask what a concept having that role would be like, what conditions would govern its application. (Craig 1990: 2).

These two steps mirror the method of pragmatic conceptual analysis very closely. In addition, Sally Haslanger's ameliorative approach to concepts deserves to be mentioned here. It starts with the question about the purposes for which we have a concept and then asks how we should understand its meaning so that it would best realise these purposes (Haslanger 1999: 468; 2000: 34; 2005: 11; 2006: 95). Again, the similarities with Fisher (and Craig) should be quite clear. By adopting such a methodology, the present approach doesn't

---

<sup>6</sup> Of course, this doesn't mean that we can completely ignore the already existing patterns of use.

merely describe mental state attributions. The aim is also to provide an account which *explains* how the benefits of folk psychology, which we care about, are possible.

### **1.3 Outline of the thesis**

In the next chapter I will articulate some benefits that folk psychological attributions provide us with. As I do this, I will also provide an overview of the state of research on the roles played by folk psychological concepts. More particularly, the roles I will focus upon are the prediction and shaping of the minds and behaviours of other people. After having raised some doubt over the pervasiveness and stability of those attributional benefits, I will present a third role of belief and desire attributions: affectability. This involves the immediate engaged responses that attributers have to attributees. In the case of belief attributions, these responses are agreement or disagreement with the attributed belief. In the case of desire attributions, these responses are endorsement or disapproval of the attributed desire. I will argue that affectability is indispensable for folk psychology. If engaged responses weren't open to attributers, mental state attributions as such would be unrecognizable to us.

In the third chapter I will elaborate on the notion of affectability. I will articulate further responses which are based on the immediate reactions of agreement/disagreement and endorsement/disapproval. These involve adopting an attributed belief or desire oneself, having emotional reactions to an attributed attitude and negotiation with or manipulation of another's attitudes. This will be the most empirically informed chapter of the thesis. It is here that I will provide evidence for the pervasiveness of the aforementioned responses from different disciplines.

The fourth chapter turns to the second aim of this thesis: to explicate the application conditions of the concepts of belief and desire. I will argue that the best way to do that, which takes into account the indispensability of engaged responses of agreement/disagreement and endorsement/disapproval, is to understand belief and desire attributions in terms of speech acts. The best contender for this kind of account is Christopher Gauker's (2003a) communicative conception of mental states according to which an ascription of a belief is an assertion on someone else's behalf and an ascription of a desire consists of a request on someone else's behalf. I will defend this view against most immediate objections and indicate how it can also explain the other benefits of folk psychology besides affectability; such as prediction, explanation and mindshaping.

The next chapter takes a critical stance towards the communicative conception. Although the communicative conception offered an initial account of the application conditions of folk psychological concepts, it is crucially lacking in certain respects. I will argue that Gauker's proposal is unable to fully explain the need for the concepts of belief and desire because the latter are too tightly connected with vicarious speech acts. I will then try to improve the original

communicative conception by offering an explanation of why the concepts of belief and desire need to be introduced to the practice of linguistic communication. The introduction of these concepts enables one to connect requests on another's behalf with the question whether the fulfilment of those speech acts would offer subjective satisfaction (pleasure) to the person on whose behalf the request is made. The adoption of the concept of belief, on the other hand, will be explained by the need to assert something on another's behalf even if that person herself hasn't asserted it because, considering her behavioural dispositions, she should assert that. With these explanations at hand, I can finally fully explicate the application conditions of the concepts of belief and desire. Since these conditions still make reference to vicarious utterances, I take it to be a modified version of the communicative conception.

In the final two chapters I will respond to possible concerns about the limits of the modified communicative conception. In the sixth chapter I will investigate the ontological significance of the previous analysis. I will explore the various ways in which one can draw ontological inferences about beliefs and desires from facts about their attribution. Eventually I will argue that the conceptual explication that I've provided in the previous chapters allows us to construe the ontological commitments of folk psychologists in a minimalist way. The truth of attributions of belief and desire requires only that certain facts about behavioural and hedonic dispositions obtain and nothing else. In the second part of the chapter I will also address the question of whether these minimalist commitments imply interpretationism or realism about beliefs and desires and defend the second answer.

In the seventh and final chapter I will consider self-attributions of belief and desire. These have some peculiar characteristics which are not shared by third-person attributions. I will try to fend off a concern that the modified communicative conception is unable to account for these peculiarities. In the first part of the chapter I will inquire into phenomenology of self-attributions and bring out their relative immediacy. This immediacy derives from the fact that, unlike third-personal attributions which involve assertions or requests on *someone's behalf*, self-attributions consist of (wholehearted) assertions and requests *simpliciter*. In the second half of the chapter, I will address the question of whether self-attributions are in some way epistemically privileged and defend the view that there is such privilege and that this privilege reduces to the contingent fact that a person has more information about her behavioural and hedonic dispositions than other people.

## **I.4 Concluding remarks**

The present thesis focuses on an aspect of beliefs and desires which hasn't been the centre of attention in philosophy of mind. Other concerns have been more prevalent. For instance, one of the central projects in philosophy of mind concerning intentional mental states such as beliefs and desires has been to

naturalize mental content (Dretske 1981; Fodor 1987; Millikan 1984). There have been vigorous attempts to explain content in terms of more basic and more scientifically respectable relations and properties. In this thesis I won't take a stand on this issue. My acceptance of the idea that belief and desire attributions can be understood in terms of speech acts might leave the impression that I am committed to understanding mental content in terms of linguistic meaning.<sup>7</sup> This impression is wrong. I don't think that any reduction of intentionality to linguistic meaning has to be involved here and I also hint at some considerations in support of this thought in the fourth chapter.

Since it seems that the program of naturalising content has somewhat run out of steam in recent years (Godfrey-Smith 2006), it is reasonable to ask different questions about intentional states and conceptualise them from new angles. In this thesis I aim at illuminating an aspect of beliefs and desires – affectability – which is indispensable to their intelligibility to us. Since this aspect is at least as important for understanding those attitudes as the nature of their content is, its study is philosophically well-motivated. This doesn't mean that the more standard questions about beliefs and desires shouldn't be asked. But a fresh look at what we do when we ascribe intentional states to one another might also force us to see anew what the *explanandum* of a theory of content should be like.

Another limitation of this thesis is that it doesn't deal with other elements of the mind besides beliefs and desires. Its central idea – affectability – characterizes only these attitudes. This is why other mental states such as sensations, emotions, mere thoughts or intentions will be left out of the purview of my account. But this should be a virtue of my approach because a treatment of mental phenomena in general would lose sight of the various differences between these phenomena.

Finally, a challenge to the present approach may take an empirical form. If the question is what we do with certain concepts, the result should be an empirical theory (Price 1992: 400). One may ask, why this task couldn't simply be deferred to psychology and/or anthropology? What can philosophy contribute to solving that question? My first response is that the present approach involves an evaluative element. It is the question of *why we should* care about what we do with mental state concepts that distinguishes this approach from empirical theories. But in addition I must also stress that the picture of beliefs and desires that I develop in this thesis will not be totally cut off from empirical work (which becomes especially apparent in the third chapter). It just engages with beliefs and desires from a more general perspective and tries to gain an eagle-eye view over a certain area of mentalistic discourse. In the end, its claims can still be brought under empirical scrutiny. Nevertheless, it is still possible to judge the acceptability of these claims without such a scrutiny. As long as it is possible to evaluate philosophical propositions by giving arguments for or against them, this exercise isn't worthless. In any case, the present approach should be judged by its fruits.

---

<sup>7</sup> It is an open question whether this would amount to the naturalization of content.

## **2.THE IMPORTANCE OF AFFECTABILITY**

### **2.1 Introduction**

In this chapter, I will set the scene by introducing the basic notion of this thesis, which I call “affectability”. The role of mental state attributions is something that has been discussed in philosophy of mind to quite a large extent. Nevertheless, I will try to demonstrate that the idea of affectability is something new and isn’t well captured well by present models. Concisely put, my claim is that there are specific ways of being affected by belief- and desire-attributions that need to be available to attributers in order for the whole attributional practice to make sense. By that I don’t mean to imply that affectability captures the whole significance of these attributions. The present chapter is more about situating affectability within a larger landscape than it is about opposing the idea to other features on that terrain. Most importantly, I claim that affectability can’t be reduced to the other roles that mental state attributions play. It’s one of the basic elements of our folk psychological practice. What’s more, the responses to other people that are enabled by the affectability of mental state attributions are more stable benefits than those that have been implied by other models. But before presenting my argument, we should consider what other models of folk psychology’s function are on offer.

I will start in §2.2 with what I consider to be the standard picture according to which the basic benefits that we gain from mental state attributions are explanations and predictions. I make a distinction between these two types of benefit and concentrate on the latter. After expressing some doubts about the centrality of the predictive role in §2.3, I consider the mindshaping hypothesis as an alternative in §2.4. Having done that, in §2.5 I finally articulate the notion of affectability and in §2.6 defend its independence from other benefits of belief- and desire-attribution, and also its centrality for our folk psychological practice.

### **2.2 The standard picture**

The idea that folk psychology is an explanatory and predictive tool is quite widespread in contemporary philosophy of mind (Segal 2000: 37; Stone & Davies 1996: 120; Stich & Nichols 1993). It is actually rather tedious to bring out the references to authors who have relied on this assumption, since this idea can also be quite implicit, and even if it is brought out, there aren’t usually any arguments given to support that assumption. Nevertheless it clearly needs serious consideration because propositional attitude terms do figure in speech acts that are either meant to explain the behaviour of persons who are characterized with those terms, or provide an opportunity to anticipate their further actions. Just to give a trivial sounding example, we can say of John that he went to the store because he wanted to buy some milk and believed that the store provided that. Or, if we know that John wants to buy milk and believes



that the store sells it we can presume that he will go to the store, at least if he doesn't have any countervailing desires or beliefs. Belief-desire explanations and predictions seem to be a standard fare, or so it at least seems.

Things are a bit more complicated than that. The previous examples may give an impression that folk psychological explanations and predictions are symmetrically related, so that whenever we explain someone's behaviour in terms of beliefs and desires we could have also predicted this behaviour if we had known her attitudes, and that whenever we make a successful folk psychological prediction of behaviour this can also later be explained by those beliefs and desires that figured in the prediction. If that assumption holds, we could consider explanatory and predictive functions of folk psychology together, so that they can be simply subsumed under one role, labelled as "theoretical".

The assumption of symmetry has received some criticism from Kristin Andrews who has argued that since explanations are context-dependent (a feature owed to their being answers to specific why-questions<sup>8</sup>), there are cases when it is inappropriate to cite certain beliefs and desires when explaining someone's actions while the same attitudes were relied on in predicting those actions. Her case concerns John who hides in the closet when there is a burglar in the house. The explanation of his behaviour in terms of his belief that there's a burglar in the house and his desire not to be shot isn't appropriate when the question is "Why did John hide rather than attack the intruder?" or "Why did John avoid the burglar rather than run out of the house?" We may be interested in why he has such a belief and such a desire in the first place and in that case simply citing them in explanations isn't satisfactory. Yet it is presumably possible to rely on those attitudes in predicting that he will hide in the closet (Andrews 2003: 214). She also argues that there are ways of predicting behaviour that are often successful, predictions that rely on statistical induction, for example, but which are quite irrelevant and insufficient as explanations (ibid. 215). It isn't usually satisfactory to explain someone's behaviour by saying that she usually just acts that way but induction is often useful for predicting. Predictions seem to be in general less demanding than explanations, the only important thing being that the behaviour that what is predicted will come about, whatever the inferential steps are, while the success of an explanation hangs on the specific why-question that is being considered.

What kind of why-question is appropriate for explanations in terms of mental states? When we are interested in folk psychological explanations that involve propositional attitudes, one of the most frequently discussed criteria of evaluation is that these explanations also justify behaviour by indicating why it is rational to behave that way (Davidson 1963). The explanation of why John went to the store rationalizes his behaviour by indicating that a reasonable way to satisfy a desire for milk is to go to a store and buy some; hiding in a closet when a burglar is in the house is also quite reasonable if one wants to be safe

---

<sup>8</sup> This idea about the pragmatic nature of explanations goes back to van Fraassen (1981).

from harm.<sup>9</sup> Physiological or biological explanations don't seem to require the same kind of rationalizing as folk psychology does, which makes the latter quite distinctive – we are sometimes just dissatisfied with other kinds of explanations, for example, when we want to be able to criticize the person whose actions we explain. It isn't clear that the same applies to folk psychological predictions. Since the sole aim of predictions is to give us accurate information about future events, however that prediction is arrived at, one need not be troubled with whether we predict someone's behaviour folk psychologically, or not, as long as the predicted behaviour will come about.

Someone may object that I operate here with an overly stripped-down notion of prediction while I speak about explanations in a thick sense, when predictions may actually involve an attempt to criticize the way that the person whose behaviour is being predicted deliberates. This hopefully won't turn out to be simply a terminological issue but I'm sceptical about this objection because I don't think that it is necessary to include such concerns about the activity of making predictions if our only concern is to arrive at the state of affairs that is being predicted. If predicting behaviour also involved criticism then this kind of interaction wouldn't just be prediction. It would be prediction plus criticism. On the other hand, to provide an explanation which points out that the reasoning behind certain behaviour was unjustified is already to criticize it. This is largely an intuitive argument though. The most I can hope for is that this reasoning suggests that the asymmetry between explanations and predictions is real and noteworthy, and they can be kept apart when discussing the roles of mental state terms in our linguistic practice. I drop, therefore, the notion of theoretical function as a possible term that could unify the predictive and explanatory functions under one concept. I take them as separable.

I would express this separability such that predictions form the forward-looking face of folk psychology and explanations form the backward-looking face. This means that, in making predictions, the attributer is oriented toward future events in relation to the attribution, and to past events in case of explanations. In the context of this chapter, the focus is on the forward-looking face of folk psychology because the notion of affectability also characterizes the forward-looking face. It thus contrasts better with the predictive role – after all, if I am affected by the attributions, my reactions are oriented toward the future. It means that I take a stand on how to relate to the person who I've already interpreted, among which one option is to predict what she will do next. In what follows, I argue that the predictions that we gain from mental state attributions are less reliable than the standard model seems to assume. This paves a way to asking what else does folk psychology offer to us besides and before predictions in order for us to be able to characterize its forward-looking face more comprehensively.

As for folk psychological explanations, it is quite clear that explaining others' behaviour in terms of beliefs and desires provides an important

---

<sup>9</sup> It is arguable, of course, if there aren't any better means to keep oneself safe from burglars.

intersubjective benefit. It enables us to rationalize the behaviour in question. The attribution of a desire provides information about what the agent considers worthy of pursuit and the attribution of a belief helps us to see how the agent represented the means to satisfy that desire (Davidson 1963).<sup>10</sup> Folk psychological explanations thus inform us about the subjective reasons of why the agent acted in the way she did (Malle 2004). A full characterization of folk psychology has to take belief- and desire-explanations into account. The following discussion isn't meant to deny that. I come back to the explanatory role of mental state attributions in later chapters, where I show how it relates to affectability in interesting ways. But the focus of the present chapter is on the forward-looking face of mental state attributions.

### **2.3 On the predictive role of folk psychology**

Prediction is certainly a prominent function that is stressed when philosophers and psychologists talk and write about folk psychology. Even the idea that folk psychology is first and foremost a predictive tool is quite widespread (Carruthers 1996: 28; Dennett 1991b: 29; Lahav 1992: 104; Stich and Nichols 2003: 60). It is true that the predictive role of folk psychology needs serious consideration because propositional attitude attributions do seem to play an important role in enabling us to anticipate the further actions of those who have them.

One way to argue against the centrality of the predictive function of folk psychology is to simply emphasize that people's behaviour can be predicted by other means; for instance, by straight induction or by relying on the linguistic competence of individuals (Gauker 2003b), or by invoking norms which individuals are meant to conform to (Andrews 2009). I do not think this is sufficient to problematize the predictive role of folk psychology, though, because one could stress that, despite the alternative ways of predicting the behaviour of others, mental state attribution provides an additional way of making predictions and that is enough to demonstrate that folk psychological practice is a predictive endeavour.

Instead, I hope to show that prediction is a rather unstable good of folk psychological attributions. Surely, being able to anticipate the behaviour of entities in the environment is, without a doubt, a crucial prerequisite for successful coping and even survival. It is an asset that is never too much to be had. It is also noticeable how almost any kind of concept, unless it is very abstract and not related to the empirical world, helps us anticipate the behaviour of its referents. Since possessing a concept usually involves knowledge about causal relations and dispositions that its referents have to other objects, it is usually the case that, when applying a concept in a specific situation, we can

---

<sup>10</sup> By admitting this, I am not claiming that this is the only possible form of action-explanations.

also form certain expectations concerning what can and cannot happen<sup>11</sup>. To argue for that assumption it is not necessary to be a theory theorist about concepts or to equate concepts with abilities. All I am saying is that concepts usually enable us to predict what will happen and I take it as rather unproblematic. Even terms like “chair” and “stone” are fit to be proper examples here: if I identify something as a chair, I can expect that people may sit on it, if I identify something as a stone, I can presume that it hurts to hit your leg against it and so on.

The point of these trivial-sounding examples is that the predictive function discriminates poorly between different vocabularies, unless it is explained what is specific about folk psychological predictions. If someone only states that we use mental state terms in predicting behaviour, this is as uninformative as it can get. It is still possible to argue, for example, that folk psychological predictions are distinctive because they appeal to intentional states. The idea would be that they work on the presumption that people are sensitive to the content of the attributed attitudes; for instance, that they try to bring about states of affairs that satisfy their desires with corresponding contents. This is quite vague, although one can probably articulate it in such a way that the distinctive aspect of folk psychological predictions can really be brought out. In the present context, I’d prefer to consider an aspect of it which rather indicates that (*pace* Lahav 1992, for instance) folk psychology is actually quite a poor device for making predictions. This is stronger than Andrews’s or Gauker’s claim that there are alternative ways of predicting behaviour.

By folk psychology’s predictive poverty I mean that, at least in everyday cases, mental state attributions don’t really give us much assurance over the way in which the person, who is being interpreted, will behave. First, there is the simple phenomenological observation that it isn’t usually possible to predict with certainty that people act in accordance with their beliefs and desires. Just consider everyday cases when we are aware that others want and believe something. Even if this knowledge can inform our expectations, we also wouldn’t be surprised when the other didn’t act in accordance with those attitudes (cf. Morton 1996). The attribution of mental states can have satisfying results independently of our ability to use those states to predict what will happen next. Even the simple example of John wanting milk and believing that the nearby shop provides it doesn’t really illustrate any predictive certainty concerning his future behaviour. This becomes even more apparent in complex, multi-agent situations.

One could form a predictive law, in the form of a practical syllogism, which can be expressed thus: if a person wants that *p*, and believes that doing X will bring about *p*, he will, all other things being equal, do X (cf. Gopnik & Meltzoff 1997: 126). The *ceteris paribus* (all other things being equal) clause is crucial here because otherwise this law would turn out to be wrong most of the time.

---

<sup>11</sup> A possible exception here could be phenomenal concepts the application of which supposedly doesn’t involve any inferences concerning their causes or effects.

There are always numerous factors that affect people's actions. This isn't as such an objection to the idea that this kind of predictive law enables us to successfully predict behaviour of persons; one can simply point out how *ceteris paribus* clauses inform laws in special sciences also (Fodor 1987). Still, the strong dependency of the applicability of those putative predictive laws on the circumstances of application, which can be extremely complex, may throw some doubt on the idea that prediction is somehow the most stable benefit of folk psychological attributions. The problem is exacerbated if we take into consideration the holistic character of folk psychological attributions: an inference from attributed belief-desire pairs to subsequent behaviour needs to take into account other possible beliefs and desires the attributee might have (Zawidzki 2008). These are well-known troubles with the simple predictive model which different authors have tried to resolve, by different means, by setting different constraints on possible interpretations, proposing that mental state attribution relies: entirely (Goldman 1989) or partially (Goldman 2006; Stich & Nichols 2003) on simulation; on prior situational and contextual knowledge (Wikerson 2001: 141); on narrative practice (Hutto 2008); on the understanding of social roles (Bermúdez 2003b). Finally, Zawidzki himself has suggested that successful and reliable mental state attribution requires prior mindshaping. But the need to complement the model indicates that predicting merely on the basis of beliefs and desires is a difficult feat to pull through.<sup>12</sup>

Another consideration for thinking that folk psychology's predictive potential in itself is quite poor, concerns rationality. There are authors, most notably Davidson and Dennett, who have claimed that folk psychological interpretations are based on the assumption that the objects of interpretation are rational – that they don't usually believe contradictory propositions, that they choose the best means to their ends, follow their beliefs to their logical conclusion etc.<sup>13</sup> This assumption is very much contested mostly because of its relative vagueness. Still, one aspect of it deserves to be brought out in the present context. John Searle has proposed that it belongs to our conception of rationality that rational agents are not directly moved by their beliefs and desires. Instead, they are able to stand back and evaluate them. Merely being caused by beliefs and desires to behave isn't sufficient for rationality; rather, that kind of behaviour is nonrational (Searle 2001: 12)<sup>14</sup>. I am not sure how much here hangs on the notion of rationality as such. Searle's idea is only meant to illustrate how we often expect persons to be able to stand back and reflect on

---

<sup>12</sup> What's more, in everyday situations, one also needs to reckon with the emotional and affective states of the attributees, which also influence how they will eventually act (Andrews 2012: 53).

<sup>13</sup> These were just some possible examples of rationality requirements, not attributed to specific authors.

<sup>14</sup> A bit similar idea according to which persons should be able to take a second-order perspective towards their desires and not simply act on whatever first-order desire they may have, can be found from Frankfurt (1971). See also Velleman (1992), about the way how classical belief-desire explanations leave agents out of the picture.

their desires and beliefs. If this kind of expectation has its place in our linguistic practice, the attribution of those states, then, shouldn't be directly predictive, at least not in the sense that the attribution of inner states serves to predict behaviour in science. In order to honour the rationality of the other, one must reckon with the fact that she may not act in accordance with her beliefs and desires, and it shouldn't be surprising when she doesn't.

I also tentatively claim that the attribution of propositional attitudes may make others actually more unpredictable than they were before the attribution. Let's assume that we have regularly predicted, on the basis of induction, that when John goes to a restaurant he chooses a table next to a window. The prediction works just fine and we assume that this is simply due to John's habit of picking tables. Contrast it with the case when we know that John wants to see what happens on the street while eating and believes that in order to see that, one needs to sit beside a window. Now we must reckon with the possibility that John can evaluate his beliefs and desires, change them, decide whether to act on the basis of them, and even be irrational and not notice any connection between his attitudes. Inductive generalization, as a rigid method, seems to be relatively more trustworthy as long as one *can* rely on it. Bringing in propositional attitudes complicates the matter and adds a dose of unpredictability because the interpreted person can presumably decide whether or not she will act in accordance with her attitudes. I admit that the claim of unpredictability is only tentative because it depends on the theory of practical reasoning one is willing to adopt, and if that theory relies on a belief-desire model of decision-making, its supporter could argue that, if our attitude-ascriptions enable predictability, we should just look further for the really motivating beliefs and desires. But it makes sense to ask why we even need to look further if maybe there is nothing to be found. Actually, the idea that folk psychological attributions might impede our ability to make accurate predictions has explicitly been noted recently by Andrews (2012: 107). She appeals to psychological evidence (see Wilson & LaFleur 1995) according to which people tend to fixate on their initial assumption about others' beliefs and desires and ignore or underestimate other, possibly more accurate, hypotheses. This is a somewhat different consideration from mine. My consideration appealed to the rationality of attributees but its overall point is the same: the ascription of beliefs and desires doesn't guarantee an increase in predictability.

As for the final consideration against the centrality of prediction, it is useful to consider certain strategic situations of decision where predictions merely on the basis of attitudes turn out to be intractable because agents' decisions are contingent on each other. A good illustration here is the game Stag Hunt<sup>15</sup> wherein two persons need to decide what to hunt: a hare or a stag. If both persons hunt the stag, they receive the largest benefit. If, on the other hand, only one of them tries to hunt the stag and the other hunts the hare, then the first will

---

<sup>15</sup> The Stag Hunt story was first described by Rousseau (1984 [1975], Sec. III), and given a game theoretic formulation by Skyrms (2004).

receive nothing, while the second still gets something. An attempt to predict what the other will do solely on the basis of his beliefs and desires turns out to be tricky: what the other decides to do seems to depend on what he believes I will do but what he believes I will do depends on what I believe he will do but what I believe he will do depends on what he believes I believe he will do and so on *ad infinitum* (cf. Zawidzki 2013: 109). The problem seems to be solvable when an explicit verbal promise and the assumption of trust is brought into play but that would mean that the successful prediction relies on a speech act and not an attitude. Although this issue has more limited force against predictions claim to centrality, given that much of our social life consists of such situations wherein our decisions are mutually dependent on each other, the issue still poses a considerable problem for those who want to see prediction as the most reliable good of folk psychology.

We can say, then, that the predictive role of mental state attributions is rather limited and much less reliable than many seem to assume. The predictive possibilities of folk psychology are far from stable. This gives us a lot of room to consider other possibilities of characterizing the forward-looking face of folk psychology.

## 2.4 Alternative: mindshaping

To what else might be involved in the forward-looking face of folk psychology (besides prediction), let's consider some alternative proposals. For instance, Joshua Knobe has claimed that folk psychology, unlike scientific theories, is influenced by our moral judgements (Knobe 2010). According to him, people's attributions of intentionality depend on whether or not an action's side effects are morally acceptable – if they are, people are less inclined to attribute intentionality to the action than in the case of negative side effects. If a scientist were to change her causal attributions depending on evaluative considerations this would hardly be acceptable and would instead be taken as a sign of a lack of objectivity.<sup>16</sup> Folk psychological attributions lack that kind of ideal. Knobe's idea concerning the so-called “Knobe effect” has been under heavy theoretical and experimental scrutiny in recent years which indicates that its status is one of a falsifiable hypothesis.

However, here I am going to consider a different view which at least on a superficial level shares with Knobe's hypothesis the idea that folk psychological attributions differ in their function from scientific explanations and predictions. According to it, the attributions function as a mindshaping device by which we mould people's behaviour and make normative demands on them. Although perhaps less precise than Knobe's claim, the mindshaping hypothesis is also more pretentious because it doesn't only claim something about factors that are taken into consideration in mental state attributions, but also can be taken as saying something about the aim of those attributions and the general role they

---

<sup>16</sup> The case of social sciences is perhaps more ambiguous.

play in human social life. In what follows, I will analyse the work of three authors who can be associated with the mindshaping hypothesis – Matteo Mameli, Tori McGeer and Tadeusz Zawidzki – and ask what sense could be made of it. In the present context, the question is whether this hypothesis gives us any sense of the stable benefits that folk psychology offers.

Perhaps the most clear-cut presentation of the idea of mindshaping can be found in the work of Mameli who uses it to make sense of the evolution of social cognition. According to him, folk psychology or mindreading is a delicate form of human niche construction (Mameli 2001: 599). The latter denotes such practices by which organisms transform their environment and so alter the selection pressures on themselves and their descendants (Laland & Sterelny 2006; Sterelny 2007). Examples of these practices would include dam-building by beavers, nest-building by ants and termites, web-spinning by spiders, badgers digging burrows, etc. By engaging in niche construction, organisms are able to canalise their evolution, at least to some extent, not merely responding to the already existing environment which they themselves didn't create. If mental state attribution were a form of niche construction through mindshaping, this would mean that folk psychology alters or has altered the social environment of humans and these alterations should also be reflected in our genetic make-up.

What kind of effect could both be produced by mindreading skills and be a reliably recurring developmental resource? One rather obvious and already indicated response is that one of the recurring fruits of folk psychology which explains its selection is its use in predictions of behaviour. Mameli thinks that this view should be complemented by taking into account mindshaping effects. These involve any kind of modification in the mental life of a subject, brought about by attributing traits, mental states or actions to her (Mameli 2001: 608). These include such effects as people actually adopting a mental state that is attributed to them, even if they haven't originally possessed it, parents shaping the social expectations of their children by reading into their behaviour communicative intentions, people convincing themselves into having certain attitudes they didn't have before, etc. Mameli calls them also "expectancy effects" since repeated attributions of a mental state produce the state in question as expected by the attributer, and labels such attributions "expectancies" (ibid. 609). Conceived this way, expectancy effects could be taken as self-fulfilling prophecies, postulated in social psychology.

As Mameli himself is ready to point out, much of the work in social psychology can be interpreted as the study of mindshaping or expectancy effects. This interpretation is quite conceivable when it comes to stereotype attribution. For instance, gender stereotypes are at work when the results of a maths test differ between men and women because this difference may result from women having heard the stereotype of women being worse in maths than men (Spencer, Steele & Quinn 1999; Inzlicht & Ben-Zeev 2000). This indicates that the attribution of a gender stereotype to someone can bring about actions corresponding to that stereotype. As for another example, there is some evidence that entertaining cultural scripts of depression can generate depression,



so that this mental condition may at least to some extent be the product of attribution and not of any natural causes (Prinz 2012a: 284). Also, there is evidence that the idea of being an extrovert also brings about extroverted behaviour (Furnham 1982). More generally, if someone thinks about oneself or others in terms of certain traits or attitudes, her behaviour reflects those thoughts and makes them true, even if the person in question originally didn't have these traits or attitudes. The process either involves the internalization of the attributed attitudes into subject's self-conception or it proceeds more automatically by generating the expectation-matching behaviour (Mameli 2001: 614). As I understand this difference, in the first case the subject also ascribes the attributed attitudes to herself and in the second case she is unaware of forming such attitudes.

Nothing in Mameli's account, though, indicates that folk psychology, the ability to ascribe beliefs and desires, itself was selected because of mindshaping effects, although the latter may have played some part in altering our mindreading dispositions (2001: 621). The latter are still presupposed in some form and the predictive role of folk psychology comes before mindshaping. In the current context, though, where I only intend to bring out and compare different benefits of mental state attributions, the relationship of priority between the predictive and mindshaping role isn't that important. In any case, Mameli doesn't see any opposition between the mindreading and mindshaping model. According to him, folk psychology is selected for both functions (Mameli 2001: 608). Although mindreading seems to be the primary *rationale* for the selection and mindshaping is more like the icing on the cake. It should also be noted, though, that Mameli himself doesn't offer any examples of mindshaping through the attribution of beliefs and desires. His focus is not on propositional attitudes.

Tad Zawidzki's proposals are potentially more radical because he also toys with the idea that folk psychology as such was actually selected for the mindshaping function while the predictive fruits of folk psychology are quite meagre. As a result, he also treats mindreading and mindshaping hypotheses as competitors, while Mameli doesn't see any tension between them. For Zawidzki, if people attribute mental states to others they act not like scientists who are first and foremost interested in explaining and predicting behaviour but rather as legislators who attempt to make people more controllable and predictable by setting normative constraints on them using mental state attributions (Zawidzki 2008: 206). The latter function should have primacy before the predictive one which, as far as it exists, is enabled by mindshaping.

Much of his argumentation, as I see it, focuses on making the negative case against the idea that the primary function of folk psychology is to predict behaviour. These arguments aim to bring out the unreliability of folk psychological predictions which don't presuppose prior mindshaping (2008: 196), and the normative character of mental state attributions which isn't captured by the mere mindreading model (ibid. 199). I have already discussed the unreliability of folk psychological predictions (see §2.2) and I largely agree

with Zawidzki. He suggests that successful and reliable mental state attribution requires prior mindshaping. How can this work? Zawidzki uses an analogy with traffic: although it would probably be intractable to calculate how other drivers might behave if the only method for doing it were to speculate about the inner causes – beliefs and desires – behind their behaviour. There simply isn't enough time to do that when the situation requires an immediate decision. Although the problem of coordination seems to be resolvable if mindreading were the only possibility, the adoption of traffic laws provides an immediate solution to the problem. Instead of relying on individual mindreading resources in order to predict the behaviour of fellow traffickers, individuals can simply assume that others behave largely in accordance with the law (Zawidzki 2008: 199). If folk psychological attributions function analogously to normative requirements, one may hope to base a case for mindshaping on the ground that folk psychological attributions enable us to make coordination with others possible.<sup>17</sup>

As already pointed out, folk psychological explanations can be seen as rationalizing explanations, so that they make sense of people's actions in terms of the reasonableness of the deliberative process that underlies them (Bruner 1990: 47). Zawidzki seems to use that idea in bringing out how in many cases, when we ascribe certain beliefs and desires to another person and predict her behaviour on that assumption, the failure of the prediction doesn't necessarily motivate us to change or drop our attribution but rather it motivates us to challenge her behaviour as inappropriate and perhaps even sanction her for not conforming to the norm (Zawidzki 2008: 201). These sanctions don't just arise in the case of behavioural predictions. They arise elsewhere too. For instance, if we know that the fact that *p* holds, we may demand that another person believe that *p*, even if the attribution of that belief seems to be false, and so we try to make our attribution true by convincing her that she should believe that *p*. Or, in case of desires, we might require others to have certain desires, even if our original attributions don't correspond to the facts. These putative aspects of mental state attribution, then, seem to indicate that folk psychology isn't merely an epistemic practice but involves a political element. As argued for in this way, mindshaping hypothesis may receive much more direct support. But to see how this aspect of folk psychology could be elaborated, I turn to Victoria McGeer's account.

McGeer doesn't explicitly talk about mindshaping but her account of the regulative dimension of folk psychology basically expresses the same observations concerning the way that we often don't revise our mental state attributions after predictive or other kinds of epistemic failure. Rather, we sanction or challenge the attributee for not conforming to the attribution in question (McGeer 2007: 148). In the end, she is much more modest than

---

<sup>17</sup> In his recent book on mindshaping, Zawidzki (2013) doesn't seem to take the function of folk psychological attributions to be mindshaping as such; rather, they function to rationalize counternormative behavior. This is a view that is advocated also by Hutto (2008) and Andrews (2012). The nonmentalistic mindshaping practices are supposed to make the attribution of beliefs and desires possible, though.

Zawidzki because she doesn't seem to identify the mindshaping role with folk psychology's selected function. McGeer's proposals, then, are quite in line with the weaker version of the mindshaping thesis according to which the latter is meant to bring out a certain normative aspect of our folk psychological practice which can be taken as valuable and which has to belong to a comprehensive account of folk psychology.<sup>18</sup> It is noticeable that she doesn't concern herself with the question of function that folk psychology is meant to fulfil. Still, McGeer intends her thesis to have developmental consequences and thinks that it could reconceptualize the idea of folk psychology as a regulative practice.

Drawing inspiration from Sellars (1997), she claims that the attribution of belief or knowledge to oneself pushes us to live up to the attribution, or if others attribute such states to us we can either challenge the ascription or try to fulfil those expectations which the attitude in question amounts to (McGeer 2007: 146). Mental state attributions, then, set certain demands on the attributers and attributees which they try to answer by constraining and canalizing their behaviour in certain ways. One of the advantages of this view of folk psychology is similar to Zawidzki's point, namely, that regulation eases the interpretive burden by constraining the range of attributions and their behavioural implications. This suggests that folk psychological understanding takes place in an intersubjective give-and-take where participants can work together to form an appropriate interpretation of one another (ibid. 149). I am not sure about this claim because although reciprocal interaction may reduce that range of possible interpretations, compared of a detached standpoint, this point seems to be independent from the thesis about folk psychology's regulative dimension – also an *opponent* of the mindshaping hypothesis could accept that reciprocal interactions enable people to constrain their interpretations by improving their access to relevant information.

Another similar point with Zawidzki that McGeer makes is that the mindshaping hypothesis makes folk psychological interpretations more fluent by off-loading the relevant knowledge into the external environment – since the latter is shaped by folk psychological and other norms, the interpreters don't need to rely on some rich internal knowledge-base but instead can simply exploit the knowledge of our shared social practice (ibid. 150). What's more – and this is more of an idiosyncratic aspect of McGeer's proposal – this knowledge is a form of know-how, as opposed to know-that. This should make interpretational understanding less mysterious (ibid. 151) but actually it is quite obscure what exactly that kind of know-how amounts to and McGeer doesn't provide much clarification in that regard.

As for developmental implications, since the mastery of folk psychology requires involvement in the regulative practice of one's community, the social

---

<sup>18</sup> “Even on the standard view, then, folk psychology is not just an explanatory/ predictive practice; it is also, in a sense, a normative practice: a practice of showing how people's performances live up to certain norms and thereby become, in that special way, intelligible.” (McGeer 2007: 141)

experience of children is ordered in the way that they get introduced to structured interactions wherein they learn to respond appropriately to others both by forming the right mental states and the correct behaviour, thus mastering the norms of folk psychology (ibid. 153). Since McGeer stressed the mastery of folk psychology to be a form of know-how, the development of a child probably shouldn't be seen as improving a grasp on propositionally articulated folk psychological principles but instead as a development of the relevant know-how which consists of responding properly to normative requirements. This idea can find some empirical support from the plausible claim that the mastery of norms precedes the folk psychological understanding in children's development (Andrews 2009), although the latter claim doesn't necessitate the former. I suspect that the theoretical advantages of the regulative view that McGeer proposes wouldn't convince her opponents. As already noted, her account of the know-how that is required for competent folk psychologists is quite rudimentary. Also, although the idea that the regulative dimension of folk psychology can act as a constraining effect on possible mental state attributions has some plausibility, the description of that process is left somewhat implicit. Still, these possible gripes concern the fact that this account isn't yet fully formed which means that it requires elaboration, not rejection.

## 2.5 Evaluating the mindshaping hypothesis

What to make of the mindshaping hypothesis, then? Can it challenge any fundamental assumption in the standard view of folk psychology? By "the standard view" I mean the relatively loose yet still tractable idea, brought out in the beginning of the chapter, that when we ascribe mental states to ourselves and to other people, the main forward-looking benefit is that those attributions providing us with an opportunity to predict behaviour. It *is* a challenge to that kind of model, in the sense that it points out that there is an aspect of folk psychology that isn't accounted for. Although conceived on a very general level, without many direct empirical consequences, it can be used as a critique against the aforementioned philosophical assumption about folk psychology and it hopefully discourages philosophers from uncritically claiming seeming trivialities such as "*the* function of folk psychology is to predict behaviour". In the present context, it is sufficient to confine ourselves to the interpretation of the mindshaping hypothesis according to which mindshaping is one of the forward-looking benefits that folk psychology has, along with prediction. I won't address the question whether it has any implications for development, although it probably does.

Having brought out two elements of the forward-looking face of folk psychology, prediction and mindshaping, it is finally time to introduce the notion of affectability. In order to arrive there, we should notice that both the mindshaping and the standard view share the same focus, namely, the person who is being interpreted. In both cases the stress is on how to understand or

change her mind and behaviour. What has been somewhat overlooked is the interpreter herself and how her attributions weigh on her. We need to keep in mind that third-person mental state attributions aren't necessarily taken into account by the attributee, so that the mindshaping effects on the latter are contingent on the particular context of interpretation: on the motivational and cognitive states of the individuals involved and also on the material and social environment. For instance, the other person might not simply care about the sanctions that the interpreter wants to force on her. There might be misunderstandings between the two, or there might be something wrong with the information channels between them. There are multiple ways in which the mindshaping process can go awry and result in a failure to shape the mind of the attributee. This fragility is comparable to the unreliability of folk psychological predictions which I brought out in §2.2. Both in cases of predicting and mindshaping there is a high level uncertainty with regard to those endeavors actually succeeding.

The attributer herself, on the other hand, is always affected in some way by her own attributions. Assuming that she was motivated to make them in the first place, it should be the case that they bear upon her own understanding of the intersubjective situation. This means that actually the most reliable effects that the mental state attributions have might concern the attributer, not the attributee. So it is also the attributer who we should perhaps look in order to see how she is shaped by her own attributions because in her case she is affected in some way or other and the process is more direct. In that sense, it should be possible to talk about first-person mindshaping. In what follows, I intend to elaborate on that idea and show that there is a kind of stable benefit of folk psychological attributions that comes precisely from the fact that the attributer is affected by her own interpretation of others.

## **2.6 Affectability – engaged mindreading**

In order to see what other forward-looking benefits besides predictions and mindshaping are gained from mental state attributions, we should look more closely at what can happen before predictions are made. It is noticeable that one can respond to the attitudes that are attributed to another person and these responses can't be reduced to a predictive activity. Let's take some examples where John is always the attributer and Mary is the attributee. So for instance, if John thinks that Mary desires to visit Sai Baba, he might not have any specific expectation about whether she will do something in order to fulfil her desire. But nonetheless, he might (strongly) disapprove of the kind of wish Mary has. Due to that disapproval, he needs to decide, for example, whether to encourage her to fulfil her dream or persuade her to give up her desire. As an example of belief-attribution, imagine that John thinks that Mary believes that Earth was created 6000 years ago. Again, he might not have any well-formed expectations about how Mary will behave given that she has this belief but he still isn't left

unaffected by the attribution. For instance, he might disagree with it. In that case, he needs to decide whether to commend Mary, who he admires for her conviction, or instead to express his disagreement with her.

These examples might sound trivial but they express a crucial element in our folk psychological practice, which is what I mean by “affectability” or “engaged mindreading”. When we ascribe beliefs and desires to each other we ourselves are affected by our own ascriptions and this enables us to adjust our own behaviour, depending upon the content of the attitudes that we have attributed. Knowing what Mary wants and believes matters to John because it affects his possible actions – John needs to decide now whether to attempt to overturn Mary’s desire or to help her to satisfy it, and whether to persuade Mary out of her false belief or to leave her in ignorance and delusion. In these cases, the knowledge of someone’s desires and beliefs enables one to do something about it, and that is already enough to be interested in them. What others believe and desire influences us both practically and emotionally, whether or not we are able to form any well-formed predictions on the basis of those attitudes. Since both beliefs and desires have satisfaction or correctness conditions – beliefs can be true or mistaken, desires can be satisfied or frustrated – knowledge of those attitudes gives us opportunity to intervene in their possible fulfilment, whether or not the subjects of those attitudes will act on them. Note that this claim doesn’t amount to the mindshaping proposal because the focus is on the effects of the attribution to the attributer, leaving open whether changing other’s attitudes will succeed or not, or whether even moulding other’s attitudes is in the interests of the ascriber.<sup>19</sup> Instead, she might simply endorse them.

The engagedness of folk psychology is much more pervasive than this handful of examples might suggest and it isn’t limited to cases wherein the attributee has an attitude that the attributer doesn’t share. There are many different ways in which a person can be affected by her attributions to others. Nonetheless, this can be articulated quite systematically because both desire and belief have their particular response profiles. In case of attributing desires, if I attribute the desire that *p* to a person X, my first reaction to this attribution is whether to endorse or disapprove of X’s desire which I think of her as having. Depending on my reaction, I can then try to help X to achieve *p*, or prevent the corresponding state of affairs ever taking place. I may even lie to X about the possibility or impossibility of *p* or I may encourage X to satisfy her desire. So when I know that some other person wants that *p*, I can take it into account and adjust my future behaviour, depending on my own interests I have towards *p*. Although such possible actions in relation to one’s attribution are varied, I take endorsement and disapproval to be the most basic reactions one might have towards the attributed desire because the subsequent behaviour depends on those initial reactions.

As for beliefs, if I attribute the belief that *p* to X, the initial reaction is either to agree or disagree with the belief. On the basis of this reaction, then, I can

---

<sup>19</sup> However, it can be called “first-person mindshaping”.

express my agreement with her belief if I agree or try to persuade X of being mistaken in the case of disagreement. I may also form a new opinion concerning her epistemic credentials, usually positive in the case of agreement and negative if I disagree with her. In a more extreme case, I might even get offended and refuse to talk with her if our beliefs contradict each other. The attribution of beliefs might also have consequences for whether I am willing to engage with X in cooperative actions and affects the extent to which I excuse her misbehaviours. But agreement and disagreement seem to be the basic reactions which we undergo due to having attributed beliefs, thus mirroring the reactions of endorsement and disapproval in the case of desires.

In the case of desires, there's the ambiguity of whether endorsing the desire that  $p$  involves taking the content  $p$  to be desirable or simply evaluating the attitude as something acceptable. In the first case a danger might arise that endorsing another's desire that  $p$  and wanting that  $p$  oneself get conflated. After all, taking something to be desirable seems to amount to wanting it and in that case the endorsement wouldn't be a neutral act for the attributer. Because of that I interpret endorsement in the second, weaker sense. That doesn't mean, though, that the interpreter need not entertain the content of the attributed desire in order to intelligibly evaluate the attitude as something acceptable. Unlike in the case of desires, an analogous issue of ambiguity doesn't seem to arise in the case of beliefs – agreeing with a belief that  $p$  means also having (or acquiring) the belief with the same content.

Of course, between endorsement/disapproval and agreement/disagreement lies indifference. In cases where I don't really care about the attributee's desire, the question of whether I should endorse or disapprove of it need not really matter to me; and if the other's belief doesn't concern me I don't need to take a stand on whether it's true or not. But it's plausible that people mostly resort to attributing beliefs and desires when they do take interest in others' take on the world. In the end, it's an empirical question to what extent one can maintain indifference towards the attributed attitudes but I am doubtful that this can be a very stable stance. After all, the basic reactions to the other's attitudes come very cheaply. When I comprehend what the other's desire is directed at, I can already make a preliminary decision about whether the object of that desire is worth wanting or not; and if I understand the content of the attributee's belief, thus knowing its truth-conditions, I am already able to judge whether I take it to be true or not, and agreement or disagreement follows.<sup>20</sup> Even in cases of indifference towards the attributee's attitudes, then, the characteristic reactions are open to the attributer. The case of indifference still demonstrates that every actual case of mental state attribution doesn't produce an engaged response.

---

<sup>20</sup> The idea that thinking about other's beliefs and desire involves thinking about the subject matter of those attitudes is comparable to the co-cognition account of mindreading that has been developed by Jane Heal (1998). What's more, in developing her simulationist theory of co-cognition, she has also criticized the too singular focus on the predictive function of folk psychology (Heal 2003: 245). Nonetheless, she hasn't articulated the reactions of the attributer, as I attempt to do here.

Instead, the described responses are meant to turn our attention to certain possibilities.

We also should distinguish between affectability in the narrow, and in the wide sense. In the narrow sense, the engaged responses denote only the initial possible reactions of agreement/disagreement and endorsement/disapproval. In the wide sense, they characterize both the initial responses and the whole spectrum of reactions that they can bring about. In the rest of the chapter I will focus on the narrow meaning because the primary responses are like the axis around which the rest of the responses turn. Affectability, then, is something that can be quite systematically articulated, starting from the most obvious and finishing with some more speculative proposals. Although much of this idea might seem rather obvious, these ways of being affected should be kept in mind if the question is what the mental vocabulary enables us to do. In what follows, I will argue that not only can engaged responses function independently of the possibility of prediction (or mindshaping); they should also be taken as a distinguishing and necessary component of our folk psychological practice.

## **2.7 The centrality of engaged responses**

### **2.7.1 Independence from predictions and mindshaping**

Hopefully it is quite clear how being affected by someone's attitudes, at least in the narrow sense, doesn't imply being able to predict or shape her behaviour on their basis. After all, John can endorse or disapprove of Mary's desire, yet he may be quite in the dark about how Mary will behave in the future or how she may change her attitudes over time. Also, the knowledge about Mary's belief may not really help John to predict her behaviour but he can still agree or disagree with it. In that sense, the possibility of being affected by the attribution of propositional attitudes doesn't owe its existence to predictability or mindshaping and should be seen as an aspect of folk psychology that transcends those other two roles. We can always construct a situation wherein the engaged responses by means of propositional attitude ascription are still possible but no real predictions can be made and no ways of influencing the other are open. To strengthen my case, imagine a situation wherein a person's body is paralysed but she can still communicate. In that case, we are able to ascribe her beliefs and desires on the basis of her linguistic behaviour and it is also possible to be influenced by the attitudes of that paralysed person in ways as I've described. Yet we may not be able to predict or influence her behaviour, which means that the possibility of responding and other two roles, prediction and mindshaping, don't directly depend on each other. It is possible to object that we are still able to predict at least her linguistic behaviour but this already seems very restrictive because usually, when it is claimed that folk psychology enables us to predict behaviour, the latter isn't confined to speech acts. What's more, for the sake of the example, we may also stipulate that we can't predict her linguistic behaviour



either. Even in this case, the responses of agreement/disagreement and endorsement/disapproval are still open to the attributer.

It may be pointed out that the engaged responses that are independent of predictability and mindshaping can characterize particular cases but that is not really relevant when we are interested in understanding the folk psychological practice as a whole. From that perspective, the objection might go, the engaged responses are only a contingent effect of mental state attributions, derivative in relation to predictability, for instance. This objection can be refuted, though – even when considering the practice as a whole, for two reasons the engaged responses cannot be ignored: first, these responses distinguish the mentalistic discourse better from others, and second, they should be seen as a necessary element of the folk psychological practice. Let's start with the first reason.

### **2.7.2 Distinguishing the role of affectability**

Engaged responses better distinguish folk psychology from other vocabularies because predictability characterizes, to some extent, most of our conceptual machinery. Almost any concept can help us to form certain expectations concerning its referents, as long as there is some knowledge-structure that is connected with the concept in question. Being affected by the mental state attributions one makes, on the other hand, allows one to achieve something distinctive in social interactions. We can evaluate the success of a prediction on the basis of its results – whether the state of affairs being predicted occurs or not – and this evaluation is quite insensitive to the method of prediction. As long as it gives us the right result, we keep that method. If there are better means available we can drop it. If the future research in neuroscience enables us to predict the behaviour of persons by relying purely on physical vocabulary and if the predictions turn out to be accurate, as eliminativists hope, one should adopt this vocabulary in formulating one's predictions.

Being affected by the mental state attributions one makes is another matter. Compared to physical or biological vocabulary, the intentional idiom, being open to semantic evaluation, articulates different perspectives that persons can have, even if their situations are similar, that are rooted in personal histories and idiosyncratic commitments. Unlike physical objects, like chairs and stones, which can affect us only in that we must take into account possible causal relations they may enter into, the attitudes of others open the possibility of responding to them in specific ways. Since they make certain claims and demands on the world – whether how the world is, in the case of beliefs, or how it should be, in the case of desires – we can choose whether to conform or oppose them. We can also decide how to influence the behaviour of mindless objects, of course, but there is no point in negotiating with them. The latter is only possible when we acknowledge that these objects have attitudes which can be evaluated. What the engaged responses also show is how, from the first-person perspective, others' beliefs and desires influence us already due to the fact that they either correspond or conflict with our own convictions and

interests and it is up to us either to deepen the correspondence or to widen the gap. The responses of agreement/disagreement and endorsement/disapproval, then, let us see the distinctive aspects of folk psychology more clearly than its predictive aspect.

The case with mindshaping, as opposed to predictions, is a bit different because one could argue that there are specific ways of influencing others through the attribution of beliefs and desires that can't be achieved by other means. In that regard, the mindshaping thesis might be on equal grounds with the claim about affectability. I can grant that, although it seems that affectability is still more fundamental because we could imagine a folk psychological practice without possibilities for mindshaping, while engaged responses being open to the participants. But this brings us to the next point. I argue for the indispensability of affectability for the folk psychological attributions to be intelligible.

### **2.7.3 The indispensability of engaged responses**

The second reason why engaged responses can't be ignored is that these typical ways of being affected by the attributions of belief and desire should be taken as indispensable components of our folk psychological practice. This claim is rooted in the idea that if we imagine a community where the ascription of beliefs and desires functions without people ever responding to the attributed attitudes with agreement/disagreement and endorsement/disapproval, we couldn't recognize the concepts of belief and desire they use as our own. We would take it that our community and the imagined one use the same vocabulary, but the latter engages in a different conceptual activity than we do. One can't make sense of someone having a belief that can't be agreed or disagreed with, if the content of the belief is understood; and one can't make sense of someone having a desire which can't be endorsed or disapproved of, if the content of the desire is understood. Attitudes that can't be agreed or disagreed with and endorsed or disapproved of, respectively, would hardly be worth taking as beliefs and desires – these responses seem to make their use intelligible. Simply put, attributing beliefs and desires essentially involves the possibility of responding to them in pertinent ways. This may sound like a strong claim. But it is one that is hard to avoid once we have already admitted the rather commonsense assumption that beliefs and desires are attitudes to which such responses are appropriate. Crucial assumption here is that what we do by attributing beliefs and desires is experientially accessible to us, so that we should be able to recognize what is needed for the folk psychological practice to be intelligible. After all, this is *our* practice. The existence of such agential authority is by no means uncontroversial but I take it here for granted (*cf.* O'Brien 2007).

One might argue, though, that the thought experiment I presented doesn't have the force it pretends to have because without the possibility of engaged responses the practice could still include the same concepts of belief and desire as we do. The claim that they are different can only be made when concepts are

individuated in terms of their use. If concepts are individuated in terms of their causal history, for instance, then the difference in use is irrelevant. I admit that my claim about conceptual activity is ambiguous here between the two communities engaging in different practices and having different concepts. The scope of this work doesn't grant me the space to engage with the question of which theory of concepts is correct. Although I am inclined towards both interpretations of the indispensability-claim, the case can be made solely by appealing to the differences in practice. After all, I am focused here on the attribution of beliefs and desires. But those who are sympathetic to the use-based individuation of concepts should be aware of the possibility of arguing for the difference in concepts, too.

A possible worry here is that merely appealing to differences in use between two communities doesn't amount to any interesting claim about the necessary conditions for engaging in the practice of folk psychology. This is an issue about individuating practices and a claim can be made that these are social entities which don't have any specific set of necessary conditions in order to be instantiated. Lance and O'Leary-Hawthorne (1997: 191), for instance, can be taken as arguing for that claim, when they analyse the case of basketball where the so-called "Wilt Chamberlain" rules were introduced into the game, to deal with the fact that Chamberlain was too good a player. But despite the fact that the pre-Wilt and post-Wilt game had different sets of rules, they were still the same game. One might argue, then, that although the folk psychological practice with engaged responses and the practice without them are clearly quite different, they still amount to the same type of practice, and those responses need not be taken as its necessary conditions.

The Chamberlain example need not be damaging to my claim, though. It demonstrates that the individuation of a practice need not be sensitive to any change in rules that govern it but it doesn't prove that there are no components that are essential to the practice. What's more, engaged mindreading isn't articulated in terms of rules and doesn't even have any normative element built into it, which seems to indicate that the Chamberlain example might even be irrelevant. That being said, individuating a practice is still a tricky issue but I do think it is reasonable to assume that there are some consequences a practice can have that are essential to it. By "essential" I mean here those consequences the total lack of which would alienate us from what we do, so that the practice in question would become unrecognizable to us.

One might also object that the fact that people attribute beliefs and desires to nonhuman animals demonstrates that the possibility of engaged responses isn't essential to the folk psychological practice because we don't really agree or disagree with animal belief and endorse or disapprove animal desire. Note that this objection doesn't target the idea that animals don't have beliefs and desires, à la Davidson (1982). It concerns our use of the concepts of those attitudes. That makes responding to the objection somewhat easier because it is difficult to deny that higher animals have beliefs and desires. At least I don't want to enter into this debate in the present context. It is possible to argue, on the other hand,

that for our practice of attributing attitudes, attributions to animals are parasitic on more paradigmatic attributions to (adult) humans who we can respond to in the engaged manner. Although we talk quite freely of higher animals having beliefs and desires, and in certain conditions attribute those attitudes even to inanimate objects, it is still human beings who are taken as paradigmatic attributees.

What's more, this objection might not even get off the ground because it isn't completely obvious that at least the potential for engaged responses doesn't characterize such attributions. It is a separate question whether agreeing/disagreeing with animals' beliefs or endorsing/disapproving their desires makes any rational sense but if these attitudes are attributed, such responses come quite naturally to us. If that is true, that we attribute attitudes to animals might even strengthen the case for the necessity of being affected in an engaged way. It is true that when we attribute beliefs and desires to animals, we are not usually interested in agreeing/disagreeing with their beliefs and endorsing/disapproving their desires, and the predictive interests can gain the upper hand. But the lack of interest conceals the fact that these responses are still open to us. That we usually are not interested in these responses in the case of animals reinforces my impression that attributions to animals are actually parasitic on the human cases because only the latter reveal the full potential of folk psychology.

But can't similar arguments be constructed for the predictive and mindshaping roles of folk psychology? After all, it would also be hard to recognize a folk psychological practice where the attributions don't provide possibilities for predicting and mindshaping. Maybe so, but I suspect that the imagined practice without predictions or mindshaping is hard to make sense of because of instrumental considerations: it would simply seem rather pointless to have the attributional practice if it didn't have any effects on the behaviour of those to whom the attributions are made and only the responses of agreement/disagreement and endorsement/disagreement were open to us. But if the possibility of the latter were absent, not only would the practice make no sense for instrumental considerations, it would also lose intelligibility full-stop. This wouldn't be the practice of attributing beliefs and desires anymore but of some other states. Due to these considerations I think that affectability cuts deeper into the nature of folk psychology than predictability or mindshaping does. But this isn't a crucial point in the present context where the aim has been, first and foremost, to argue for the indispensability of engaged responses for understanding folk psychology.

## **2.8 Conclusion**

I haven't by any means aspired to argue that the affectability of folk psychology is the only benefit it has. I agree with Kristin Andrews that we should adopt a pluralistic view of folk psychology, so that the attribution of mental states helps us to have engaged responses to others, but also gives some predictive

advantage and can be used to mould the minds of others. Yet Andrews herself focuses too narrowly on the role of mentalistic attributions in the explanations of unexpected or counternormative behaviour, when the attributed attitudes can make sense of that (Andrews 2012, 120).<sup>21</sup> The present proposal isn't reducible to that because the need for engaged responses aren't confined to situations wherein the behaviour of another person is in some sense abnormal; she might behave as I've expected all along but the attribution of an attitude and taking an appropriate response are still beneficial to me. In addition, the advantage of having the engaged responses isn't the same as an explanatory advantage – even if the behaviour of the other has been sufficiently explained, one might still want to be able to take an evaluative stand toward her attitudes. The present view, then, doesn't merely boil down to the idea that folk psychology provides ways of accounting for abnormal behaviours.

Also, although I took the prediction and mindshaping proposals to disclose certain benefits of folk psychological attributions, the benefit I've presented doesn't boil down to them. Affectability is different from predictability for quite obvious reasons: as argued, they can be dissociated so that the engaged responses are open also in cases where predictions don't work. As for mindshaping, having the engaged responses to others' attitudes does involve a normative element, in that agreement/disagreement and endorsement/disapproval are evaluative, and thus probably at least implicitly prescriptive acts. Nevertheless, whether or not those who are having those responses aim at shaping the mind of the other is a purely contingent matter: sometimes they do, sometimes they don't. Of course, if we construe the notion of mindshaping very broadly, the engaged responses can be seen as forms of first-person mindshaping – by taking the evaluative stance on whether to agree/disagree or endorse/disapprove, one is forming and possibly shaping her own attitudes toward another person. But that isn't something that the proponents of the mindshaping hypothesis had in mind. We can conclude, then, that affectability presents itself as a distinctive and indispensable benefit of folk psychological attributions.

---

<sup>21</sup> A similar view, focusing on the evolutionary function, has been recently proposed by Zawidzki (2013, Ch. 7).

## **3. MORE ABOUT AFFECTABILITY: FROM ADOPTION TO NEGOTIATION**

### **3.1 Introduction**

In what follows, I will consider three mutually complementary ways in which the idea of affectability can be elaborated in its wide sense. The first concerns the possibility of more or less automatically adopting another's beliefs and desires after having attributed them without any explicit inference. This reaction becomes possible when disagreement or disapproval is too demanding for the person, given the social superiority of the other. Clarification of that suggestion should also do some justice to the (relatively) unreflective effects that the attributions might have. As for the second proposal, I will discuss the emotional effects that our attributions may elicit in us, drawing some parallels from Peter Strawson's seminal essay about reactive attitudes. The affective consequences of belief- and desire-attributions that are in focus there again concern the difficulty of maintaining the basic reactions that were described in the previous chapter. While the second idea is also about the unreflective effects of folk psychology, the third analysis ascends the unconscious, so to speak, and concentrates on how the attributed beliefs and desires can be negotiated so that the attributee can play an active part in the interpretation. The three parts of this chapter should thematically follow the folk psychological interpreter becoming more reflective in the process of attribution. Each of these cases shows how the initial reactions of agreement/disagreement and endorsement/disapproval can affect the following interactions between the attributer and the attributee, thus being essential to our folk psychological understanding.

So there are three additional dimensions in which the affectability of mental state attribution can be located. Although they may seem rather disparate the story will have a common theme. Besides the obvious fact that they all exemplify the idea of affectability, in all three dimensions the attitudes themselves can be in focus for the attributer and not their direct behavioural consequences. Whether in adopting a belief of the other, feeling resentment towards her desire, or trying to change her attitudes – in all those cases the represented attitude as such is the basic object of our concern. In that sense, I will retain the opposition to the predictive model of folk psychology also in the following analysis because these ways of being affected don't imply much about predictive possibilities yet still are for us something to be interested in.

## 3.2 Mimetic belief, mimetic desire

### 3.2.1 The general idea

When we are facing an attitude of another person this may sometimes require an immediate response. In certain situations that response may be simply to adopt that attitude, not necessarily being aware of one's own reaction. Emotional contagion, for instance, according to which people tend to converge emotionally by mimicking each others' behaviour, is a widely acknowledged phenomenon of social cognition (Hatfield et al. 1993, Chartrand & Bargh 1999). Emotional contagion can be taken as the basic form of automatic empathy, distinguished from the so-called controlled empathy, i.e. consciously putting oneself into another person's shoes (Hodges & Wegner 1997). People are prone to take on the experiences and feelings of one another and this may happen without much effort, by merely observing someone having that feeling. The evidence for that phenomenon comes also from neuroscience where it has been discovered that the same brain areas are activated when feeling an emotion and observing a person having that emotion, as with disgust, for instance (Wicker et al. 2003). The Humean metaphor, then, that the human minds are mirrors to one another<sup>22</sup> is alive and kicking.

Can something similar be said about propositional attitudes, such as beliefs and desires? Although I am not going to argue for an exact analogy between emotions and propositional attitudes, I will claim that one is justified in speculating that when the subject has attributed certain beliefs and desires which she disagrees with or disapproves of, respectively, she may give up some of her own attitudes in response to her attributions. There are cases when a person may start to actually believe or desire something, simply because another person is taken by her to believe or want it. The attribution of the attitude brings about having the same attitude. The directness of that response is rather analogous to the case of emotional contagion where the detection of the emotion can bring about the same emotion in the cognizer, without having to be consciously mediated. The analogy stops here, though, because the adoption of a propositional attitude from the other can't be elicited simply by behavioural cues.

Nor am I claiming that the process in question involves anything like pretend attitudes, as advocated by some simulation theorists (Currie & Ravenscroft 2002; Goldman 2006). The attitude adopted is the *same type* of mental state as the ascribed attitude. Also, the adoption doesn't necessarily imply anything about the attributer having a higher-level representation about her adopted attitude. In other words, the attributer need not be aware of even of the outcome of that process. So the process – if we only consider its starting- and endpoints – would look something like the following: first, X attributes an attitude to Y, and

---

<sup>22</sup> “The minds of men are mirrors to one another, not only because they reflect each other's emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and may decay away by insensible degrees.” (Hume, *Treatise* 2.2.5)

this motivates X to adopt that attitude herself. Of course, this can't be the whole story and now the question is what happens between those two points. If the process involved turns out to be too complicated and mediated by too many variables then it isn't clear why exactly the mental state attribution as such should be picked out as the crucial factor in attitude-adoption. For instance, if it is mediated by evidential considerations then the adoption of the attitude isn't due to the fact of previous ascription but due to rational considerations. A tricky issue surfaces when the attitude of the other person belongs to the evidential source for adopting it oneself. I'd prefer to postpone this possibility for the later occasion when I discuss the negotiability of mental state attribution. Since the more direct and presumably even irrational ways of being affected form the topic here I will be concentrating on the possibility in which the evidence doesn't come into play.<sup>23</sup>

Another desideratum at this point is to show how the attitude-adoption on the basis of folk psychological attribution differs from the one that is based on the comprehension of speech acts. When we're interested in that kind of adoption the most obvious example is the situation wherein we come to believe or desire something after hearing someone talking about it. Then one could say that the comprehension of a speech act involves inferring the attitude expressed by that utterance which is then, in certain conditions, incorporated into one's own cognitive economy. I am bringing this up because, since I tend to think that the understanding of speech acts doesn't necessarily require inferring the underlying mental states,<sup>24</sup> I distinguish it quite sharply from the attribution of attitudes that isn't necessarily tied to linguistic behaviour. The interesting question is what the latter helps us to do that is distinctive from the adoption of attitudes on the basis of speech act understanding. Then again, since in the case of beliefs most of the relevant evidence concerns the way in which people consume the (verbal) *testimony* of others, my treatment of belief-adoption is closely connected with the discoveries in that area.

So what else is needed for the process to work? I will suggest that the important variables here are the personal qualities of the attributee, most notably her trustworthiness, attractiveness, and (social) power. The overall evidence for that will be sketchy, I admit, but the present purpose is simply to point towards a way of being affected by mental state attribution that is realistic enough. In what follows, I will treat beliefs and desires separately, one reason being that the adoption of them may proceed differently and the amount of evidence that can be adduced is rather different.

---

<sup>23</sup> By saying that this process is direct I don't deny that the attribution of propositional attitudes requires semantic competence in order to evaluate the content of those attitudes. This kind of indirectness can't be avoided.

<sup>24</sup> This is a highly contested issue and due to the heavy presence of Paul Grice in the modern theories of communication the competence in communication and Theory-of-Mind abilities is often seen as going hand-in-hand (Stalnaker 1998). I come back to this issue in the next chapter.



### 3.2.2 Mimetic belief

Let's take beliefs first. If we are aware that another person believes some proposition which we ourselves haven't formed an opinion about, that proposition is brought to our attention, simply as a result of thinking about it. This also provides an opportunity to take a stand on the truth or falsity of that proposition. That much isn't contested, I hope. The question arises though: how do people come to assent to that content. On a purely rationalist picture the assent should come from considerations pertaining to whether the proposition in question is supported by evidence and how well it coheres with the ascriber's overall system of belief. The view that I support here diverges from such a rationalist view by acknowledging the possibility of non-evidential "brute" causation in the formation of beliefs. The adoption of a belief may be mediated by nothing more than the authority of the other.

A rather more radical view would claim that merely comprehending an idea carries with it an implicit acceptance, only after of which it can be called into doubt. This so-called "Spinozist" model<sup>25</sup> has been advocated by Daniel T. Gilbert, for instance (Gilbert 1991). According to him, since humans are creatures with limited time and limited cognitive processes, needing fast decisions in evidentially poor or ambiguous situations, comprehension and acceptance go together. As for experimental evidence for this thesis, it has turned out, for example, that already the comprehension of a false proposition makes it more probable that it is taken as true (Gilbert et al. 1990). If the Spinozist thesis is true, the idea that beliefs are adopted by an attributer because she has attributed those beliefs to others may lose its edge because the belief-adoption would be a too general phenomenon – as soon as some content is comprehended it is already believed. The adoption-by-attribution model isn't meant to be totally homogeneous across different contexts. It should be applicable when certain conditions concerning the attributee are satisfied.

It isn't clear, though, if the Spinozist claim is right. There is also some evidence indicating that people tend to be less liberal in assenting to a content that has relevance for them. Gilbert's experiments presented subjects with statements that weren't of any immediate relevance for them and it was less problematic to automatically accept them, while in the case of statements of high relevance the response can be more thoughtful (see Sperber et al. 2010)<sup>26</sup>. So I tentatively maintain that entertaining content is possible without accepting it beforehand. Since the primary engaged responses to the attributed belief should include agreement, disagreement, and indifference, the Spinozist model is incompatible with my model of engaged mindreading where all three are open possibilities.

What, then, is needed to bring about the adoption of beliefs that have been attributed? Since the most reasonable assumption is that one needs to be

---

<sup>25</sup> I am not doing an exegesis of Spinoza here in order to determine his actual views.

<sup>26</sup> In general, it seems that the view people have cognitive mechanisms that guarantee at least some epistemic vigilance, has recently gained prominence.

specifically motivated to end up doing that, we may also ask: what can *motivate* the attributer to make the ascribed belief her own? My guess is that the answer lies in the image the attributer has of the addressee. For example, if the latter is taken as exceptionally reliable in her belief-forming procedures, the subject may – for the sake of convenience – rely on that person by default. Signs of trustworthiness are among the first things people tend to look for in another person (Willis & Todorov 2006). After one has a reliable informant whose assertions can be trusted (Koenig et al. 2004), it is possible to trust also the beliefs that have been attributed to her. In proposing this, I am not excluding the possibility of revising the belief afterwards.

The assumption of reliability may not be the only personal quality that motivates the subject to do that, of course. In social psychology, the motivation for changing one's attitudes is quite well-studied<sup>27</sup>, and a quite useful distinction can be found from Herbert Kelman who distinguished between three major kinds of source (the person who effects the change) variables that increase the probability of being influenced by another agent - credibility, attractiveness, and means-control<sup>28</sup> (Kelman 1958: 54). All those factors may come into play in social situations when a person is aware of someone having different attitudes. If the other is a socially prominent individual or if she has a reputation of being extremely trustworthy, at least in a specific domain of expertise, or if the subject simply aims to please the addressee, the adoption of the attributed beliefs may happen quite automatically.

We should also distinguish two types of situations: one in which the ascriber already has a belief on the issue that is incompatible with the belief ascribed and one in which the ascriber hasn't yet taken a stand.<sup>29</sup> The awareness of the addressee's beliefs doesn't affect the attributer in precisely the same ways in these two cases. It is reasonable to assume that if the ascriber already has an incompatible belief then more motivational resources are needed to change it. For instance, people aren't usually inclined to change their perceptually acquired beliefs simply because another person has a diverging belief, at least not in normal circumstances (Clément 2010). The same probably applies to any kind of firmly held beliefs, although the fact that the other doesn't accept that may still affect the attributer, sometimes maybe even increasing her confidence that she must be right, simply out of obstinance, for instance. So one shouldn't expect that the adoption-by-attribution model is indiscriminately applicable to any kind of situation in which beliefs are attributed. Still, in situations where the personal qualities of the addressee satisfy the necessary requirements

---

<sup>27</sup> The notion "attitude" seems not to distinguish strictly between beliefs and desires as separate types of attitudes (Petty & Wegener 1998).

<sup>28</sup> By means-control it is meant the possibility to influence the goal-achievement of an agent (Kelman, *ibid.*). One could also call it the social power of the source, which I will do in the following discussion. I also tend to think that attractiveness could in principle be reduced to social power because, when considering its effects, it's a species of the latter.

<sup>29</sup> I disregard situations where the beliefs of the attributer and the addressee already match because one can't presumably speak about adopting any new belief on attributer's part.

(credibility, attractiveness, or social power) adoption-by-attribution presents a quick and cognitively cheap way of acquiring beliefs<sup>30</sup>.

Before closing this sub-section, I will cast a quick glance over an additional possible application of the adoption-by-attribution model. One of the most noticeable traits of human beings is our capacity for cultural learning. This means that most of the knowledge is acquired from others. Every person needs to be introduced into his community and its beliefs in order to survive in the social environment, which also brings about differences between communities because the traits acquired are specific to the group. How social transmission exactly works is an intriguing question. One form it may take could be the belief-adoption which is triggered by the attributed belief. How can this model fit into current theories of cultural evolution?

There are many theories of cultural transmission on the market. Maybe the most notorious one is the memetic approach according to which “units of culture” – ideas and behaviours – propagate in the social environment by reproducing across minds, analogously to genes (Dawkins 1976). The meme’s eye view may seem to be the best standpoint from which to defend the idea that beliefs can be adopted directly after having been attributed to them. Yet it is doubtful whether this framework can be put to work because memeticists have faced numerous criticisms over its naïve view about the way in which ideas are transmitted from one mind to another. Most importantly, cognitive anthropologists have criticized it on the ground that accurate replication occurs very rarely via an inferential mediation on the part of the receiver (Sperber 1996, 2000; Atran 2001). These authors claim that there are cognitive constraints which affect the way that the idea is processed and exclude its brute replication. Does this criticism also threaten the idea that a belief can be transmitted through adoption-by-attribution?

Of course, it doesn’t threaten the adoption-by-attribution model as such because the latter doesn’t imply anything about the similarity or difference between the attributer and the target, so the question about replication or mutation doesn’t even arise. The subject may even be mistaken in her attribution yet still adopt the attributed belief. So as long as we don’t want to give the adoption-by-attribution model any explanatory role in cultural evolution, we are safe. But I do think one can push this idea forward when we look behind cognitive constraints in belief transmission.

There are contextual biases which affect which beliefs are adopted and which are not. The explanatory importance of contextual biases has been advocated, for instance, by Joseph Henrich who has stressed especially the role of prestige in cultural transmission, among other factors (Henrich 2009). People

---

<sup>30</sup> More generally, a belief-ascription allows us to take into account another person’s take on a situation which can motivate the ascriber to consider the possibility of being mistaken, if the attributed belief diverges from hers. Rejection of belief is also a response, although merely a negative one. And, what’s most important, in the present context, is that it is possible to consider the truth or falsity of a belief independently of being able to predict anything about the attributee. It’s the attitude itself and its content that is in focus.

tend to prefer highly prestigious individuals when looking for models to learn from and the traits that are possessed by those individuals thus enjoy an advantage in survival (Henrich & Gil-White 2001). Since prestige is constituted by admirable personal qualities and is a form of social power, the adoption-by-attribution model could also find its place in the theory of cultural evolution – the hypothesis would simply be that people tend to adopt the beliefs of prestigious individuals, for the reason that *they* have those beliefs, and disregard the less fortunate.

In conclusion, the idea that merely attributing beliefs can cause the subject also to adopt them is psychologically realistic, when actualized in situations where the attributee has certain personal qualities which motivate the attributer to rely on them in forming their own beliefs. I also argued that this rather dumb and irrational, yet still cognitive, process can also play a part in cultural evolution, at least to some extent. Can something similar also be said about desires?

### 3.2.3 Mimetic desire

Assuming that acquiring beliefs by attributing them is a cognitive strategy that has at least some part to play in social cognition, what precludes us from applying this strategy to desires? The process could turn out as follows: after having attributed a desire that  $p$  to  $Y$ ,  $X$  may in the light of that fact and some kind of relevant personal quality of  $Y$  adopt that desire herself by endorsing not only the attitude but also its content. But it is unreasonable to assume that the relevant qualities in question correspond to those that effect the adoption of beliefs. At least the credibility of the attributee must be excluded from the list of source variables, simply because there is no question whether the desire is credible or not.

The credibility variable could perhaps be replaced by the indication of *practical trustworthiness*. I mean by the latter the property of having such desires and goals that conducive towards the well-being of their possessors. The well-being could be understood in a wide sense, involving simple pleasures but possibly also more complex states of affairs. If the notion of practical trustworthiness is feasible the model of desire-adoption could represent such situations where  $X$  is aware of  $Y$ 's desire that  $p$ , deems  $Y$  practically trustworthy, and because of that acquires the desire that  $p$  herself, even if the evaluation of trustworthiness is mistaken. It is doubtful whether this model is applicable to intrinsic desires, though. In order for  $X$  to take  $Y$  as practically trustworthy, the comprehension of the latter quality requires understanding what constitutes well-being for  $X$  or, in other words, what is *desirable* for  $X$ . This already implies having some relevant desires concerning one's own well-being, which are needed in order to evaluate  $Y$ 's practical trustworthiness, and  $X$  can only adopt new desires that are instrumental for satisfying those pre-existing ones.

As for attractiveness and social power one doesn't face similar difficulties. It is entirely coherent to assume that people tend to prefer things which are also wanted by attractive individuals and those who are in the position of social power. What may trouble more empirically-minded philosophers is the scarcity of evidence for the pervasiveness of such phenomena. Although the same could be said about belief-adoption, the latter idea has at least some resonance with the discoveries in social psychology about epistemic reliance upon others. Its possible role in cultural evolution is also rather dubious: while adopting beliefs after having attributed them may serve us well in exploiting the information that others have, assuming that they are epistemically reliable, the analogous method of adopting desires may easily overload the motivational mechanisms of the attributer.

The proposed idea that, in certain conditions, people tend to adopt the attributed desires isn't unprecedented, though. In literary theory, René Girard has for decades advocated a claim that the social life of human beings is largely driven by what he calls "mimetic desires". Simply put, the idea is that people tend to take over the desires that others have which often leads to potentially violent conflict, since the resources that are wanted aren't enough to satisfy all (Girard 1995). In line with my proposal, Girard doesn't think that the process of mimicking is direct; some kind of conative resonance. Instead, it is mediated by the model which includes the other person and her desire (Girard 1976) – the attribution comes first, or so I assume. Girard doesn't put much stress on the possible source variables that determine which individuals one mimics, probably because he hopes to generalize his model to most social situations. Yet one should be cautious not to over-generalize because it is doubtful whether people take over any desires of any individuals.<sup>31</sup> There are some hints of attractiveness having an effect on the process because, according to Girard, at least in early modern novels the models of desires were fictional characters out of reach, most clearly exemplified by Don Quixote whose every desire was modelled on his heroes.<sup>32</sup> Yet Girard tends to think that in the modern world mimetic mechanisms function rather indiscriminately. If he is right, the adoption-by-attribution would actually be much more prevalent in the case of desires than beliefs. Yet the question remains whether the picture offered by the literature on mimetic desire corresponds also to the psychological facts.

An attempt to demonstrate this can be found in Garrells (2006) but his actual proposals are quite sketchy and the supposed evidence he presents is actually irrelevant. The putative fact that infants are able to mimick the behaviour of other people already in the first hours of their lives (Meltzoff & Moore 1983), and that the deferred imitation also arises very early in development (Meltzoff 1988) are rather unrelated to the question about the pervasiveness of mimetic desires. He also refers to the general hypothesis that there is a common coding

---

<sup>31</sup> I may do some injustice to Girard's theory, which is actually very elaborate.

<sup>32</sup> "We literally don't know what to desire and, in order to find out, we watch the people we admire: we imitate their desire." (Girard 2004: 8)

between performed and perceived actions which enables children to understand that other people are like them. This is supported by the discovery of mirror neurons and the possible innateness of motor imitation (see Meltzoff & Decety 2003). Yet since the common coding hypothesis isn't about desires, it seems to be irrelevant in the present context.

Possibly a more interesting discovery in that respect is that infants' imitative abilities are sensitive to intentions and the goals of the person who is being imitated (Meltzoff 1995) because this involves adopting the intentions of other persons in order to fulfil their failed attempts at performing a task. Yet intentions are not desires. Imitating the first has a quite clear role to play in acquiring new skills (Wohlschläger & Bekkering 2002) while the imitation of desires probably lacks that kind of function. Also, since it isn't entirely clear whether this kind of imitation of intentions involves attributing them to the other person, it may not resonate well with the adoption-by-attribution model that is under discussion here. Finally, in the aforementioned cases, mimetic abilities can be seen as a precondition rather than the outcome of mentalistic attributions. This is implicitly brought out by Garrells himself when he points towards the hypothesis that imitation is developmentally more basic when compared with the Theory-of-Mind ability (Garrells 2006: 67).

But maybe developmental psychology isn't the best place to look for the evidence about mimetic desires, anyway, especially if we are interested in the pervasiveness of that phenomenon and what variables influence it. If we look towards neuroscience, we can find some data, according to which the activation of the ventral striatum – which is considered to be the reward area of the brain – is in positive correlation with the shared preference of a person who is considered to be an expert (Campbell-Meiklejohn et al. 2010). This seems to indicate that people are motivated (find it rewarding) to seek agreement in desires with the other if the other's opinion is valued. This suggestion remains largely a suggestion, though. I will only assume that the mimetic desires, acquired in the way described by the adoption-by-attribution model, can and do emerge in certain social situations, and that personal qualities such as social power and attractiveness (and possibly practical trustworthiness) are the main source variables that effect the process.

Minimally, from the subjective point of view, the awareness of a desire that diverges from one's own or one which hasn't been considered before, provides a possibility for considering the desirability of its content oneself. This is by no means a negligible effect of desire-attributions since adopting that desire through those considerations is a risky enterprise. But sometimes, when the attributee is respectable enough, this may actually pay off. Even if the subject is unable or unwilling to adopt the attributed desire the mere possibility of doing that can give rise to hesitation and a lack of resoluteness. It's fitting, then, to consider the emotional effects of mental state attribution next.

### 3.3 Reactive attitudes and folk psychology

One need not adopt an attributed mental state in order to be affected by it, of course. Sometimes the result is only limited to eliciting certain affective responses. It seems quite obvious that being aware of other's beliefs and desires can induce different reactions in us. Then again, this doesn't say much. One needs to specify what reactions mental state attributions produce which are specific to this capacity. That is the starting-point: even if the attributer isn't willing to adopt the ascribed attitude she usually has some sort of idea of whether the belief is something to be agreed or disagreed with, and the desire is something to be endorsed or disapproved. The affective reaction to the attitude should exhibit that kind of positive or negative valence.

The opinion that other people have certain beliefs and desires can delight or sadden us in noticeable ways. As we found earlier, a more active response would be to attempt to satisfy or frustrate a desire and to consolidate or overturn a belief. Yet the passive emotional reaction is something that probably underlies most of our attempts to do that, acting as a motivational source. One feels no motivation to effect a change in someone's beliefs if she is indifferent towards them. The same applies to desires, of course. But, for instance, if the other person wants something that I consider to be simply repulsive I probably am inclined to frustrate that desire; I certainly condemn it, at least implicitly, and this condemnation underlies the inclination to frustrate the desire in question. On the other hand, if the desire is deemed praiseworthy the appropriate reaction may be to admire the person for having that kind of attitude. Our relation to the cognitive attitudes of others is also affect-laden: another person having a false belief (or our thought of her having a false belief) can cause an attitude of condemnation in us; especially if that belief is formed on the basis of evidence that clearly points the other way. Epistemic laziness is blameworthy for many by itself, and the opposite applies to epistemic rigour. The affective reactions to beliefs aren't confined to the presumption of epistemic evaluation, though. One may have more warm feelings to another person simply because of sharing a belief with her.

The issue of our affective responses to our own mental state attributions doesn't seem to be a well-researched area in the study of social cognition. In philosophy, an analysis of an idea which is somewhat along these lines can be found in the work of Peter Strawson, whose essay "Freedom and Resentment" brought into focus the notion of "reactive attitudes" (Strawson 1974). The motivation for that essay was to re-conceptualize the question of free will and determinism, and to argue that our moral practices of ascribing responsibility don't strictly depend upon whether determinism is true or not. What matters here for my concerns is not the question whether Strawson was successful in achieving that goal. Rather, I point towards the concept of reactive attitudes in order to see whether it can be employed to account for our reactions to beliefs and desires.

Strawson draws attention to reactive attitudes to others' attitudes and intentions that are manifested in their actions; such as gratitude, resentment, forgiveness and love (1974: 4). The list is open-ended. It is meant to cover all those responses that we might have to "goodwill, its absence, or its opposite" (ibid. 6). He doesn't distinguish clearly between our responses to others' attitudes and to actions, although he stresses the first. Still, the attitudes of the other people he focuses on are those which are also expressed in action. This kind of focus on his part isn't surprising because reactive attitudes are more explicit when they are caused by others' overt behaviour. The latter may have a more noticeable effect on us than the attitude that is left unexpressed. Nonetheless, I intend to supply some considerations in favour of the idea that the awareness of others' attitudes themselves has emotional effects on the subject.<sup>33</sup>

Another thing to note about Strawson's approach is that he doesn't strictly define the mental states of others awareness of which can cause reactive attitudes in us. He speaks vaguely about attitudes and intentions (ibid. 5). Since my focus here is on beliefs and desires, it can be doubted whether the reactive attitudes considered by Strawson can even be caused by those mental states. The opposite danger would be that his account is too general and includes any kind of attitude people may have. That is why the following discussion is Strawsonian only in its inspiration but not necessarily in its content.

What kind of reactive attitudes would be appropriate to assume in case of thinking that other people believe or want something? For instance, could gratitude or resentment be such attitudes? Although some quirky scenarios may be possible, gratitude doesn't seem to be fit for that job. One feels gratitude towards another when the latter has done something that has influenced the first in some positive manner. Merely wanting or believing something doesn't merit any gratitude. The case of resentment is somewhat more intriguing. One possible meaning of resentment would be the feeling of displeasure that arises due to another agent's action that is considered to be a wrongdoing (La Caze 2001: 33).<sup>34</sup> Taken this way, the reactive attitude of resentment is first and foremost a reaction to overt behaviour and doesn't provide what we are looking for; namely, a reactive attitude to beliefs and/or desires. But resentment could be given a more specific sense. There is also resentment as *ressentiment* which is a stronger term in that it includes also the feeling of powerlessness when compared to some other person (Nietzsche [1887] 1988). *Ressentiment* arises when the subject can't do anything to influence the wrongdoer; to punish her, for instance<sup>35</sup>.

It may seem that since *ressentiment* still arises when some other person has *done* something, not simply wanted or believed something, it doesn't really

---

<sup>33</sup> I don't mean to claim that Strawson himself ignores that possibility.

<sup>34</sup> For Adam Smith, this attitude is also accompanied by the desire for revenge (Smith 1984).

<sup>35</sup> For an analysis of the difference between resentment and *ressentiment*, see Meltzer & Musolf (2002).



illuminate the possible reactive attitudes towards others' attitudes in themselves. But since *ressentiment* also involves the feeling of powerlessness, it can also be extended to others' beliefs and desires. What another person(s) believes or desires is often beyond our control, maybe even more than their behaviour, since actions can be manipulated by more means, even by adjusting the material environment (enabling conditions for action) of an agent, so that the behaviour in question simply can't get off the ground. Yet to influence others' beliefs or desires themselves is more difficult both because these attitudes aren't perceptually accessible and there doesn't seem to be any reliable method, available to a layperson, to manipulate them.<sup>36</sup> Of course, there are probably noticeable individual differences between people's abilities to influence others' beliefs and desires. *Ressentiment* can arise when a person is unable to do to that but feels, whether justly or unjustly, that the other shouldn't have that belief or desire in question that she thinks the other person has. In the case of beliefs, at least, one may in principle resort to rational argumentation to convince the other. But a person who has weaker argumentative abilities than the other may even feel *ressentiment* because of that.

As for resenting another person's belief, it is here where reactive attitudes and adoption-by-attribution, described in the previous sub-chapter, can be brought together. A person, being aware of discrepant beliefs of another person who is in the position of power is inclined to adopt those attitudes oneself but as long as she is aware of her powerlessness she also feels *ressentiment* towards the other. Inequalities between persons may actually bring about uniformity in beliefs – people in the “lower” positions adopt the attitude of the respected ones. But that uniformity can be quite unstable and artificial because of the resentfulness of the “weak”. Of course, *ressentiment* can emerge before and without adopting the beliefs of the other and the latter reaction probably can alleviate the negative attitude. After all, people tend to appreciate more those people to whom they are similar (Sally 2000).

Let's also consider the standpoint of the person in the position of power. Power over others acquires a peculiar dimension when it extends over the attitudes of the latter. Since beliefs aren't usually behaviourally manifest, don't have strongly predictable effects (as argued in the previous chapter), and can change quite unexpectedly, an attempt to control them involves a high chance of failure. In order to achieve “full” control one should also somehow guarantee that the other wouldn't feel any resentment (in the sense of *ressentiment*) while adopting the attitude because if she did, the belief that the other adopts wouldn't be fully and sincerely endorsed by her. If a person believes something and has a negative emotion because of having those mental states, it is an example of bad faith on her part because she is not able to make the attitude fully her own. She

---

<sup>36</sup> I am definitely not asserting that there are no methods. The art of demagogy, or even advertising in some cases, would be an appropriate example here. In the next section I will elaborate on that these ideas.

is alienated from her attitude.<sup>37</sup> This is why the attempt to control someone's beliefs is quite tricky – simple authority isn't enough because it creates a resentful attitude which precludes oneself from having full control over the other.<sup>38</sup> The person in power needs to control also the resentfulness of the other which otherwise may lead to forsaking that belief at the first opportunity.

But the person in control also has her own typical reactive attitudes towards those with lesser power and influence. For instance, towards those who adopt her beliefs while actually being resentful, one possible reaction would be to feel contempt or pity. If the others retain their divergent beliefs, the reaction would be harder to predict. In any case, people in opposite positions of power have both their characteristic reactive attitudes towards each other. The same probably applies to resenting someone's desires but I used the example of beliefs because the control over the desires of another person presumably doesn't imply any attempt to make her adopt one's own. However, this doesn't spell any trouble for the idea that *ressentiment* is still a reactive attitude to take towards the attributed desires of others as long as the subject feels herself to be powerless to control them.

These remarks have been hypothetical, though, and may need confirmation by research in social psychology. The aim was to provide some indication of how a reactive attitude to a mental state attribution might look. Nonetheless, I tend to think that the resentment and contempt towards other people having certain propositional attitudes is a real phenomenon. The question is rather about its extent and the underlying mechanisms<sup>39</sup>. People tend to be frustrated and feel indignation when others, whose opinions are valued, don't share their beliefs and desires, especially if they can't do anything to change them.

*Ressentiment* and contempt are only two possible reactive attitudes among many.<sup>40</sup> The feeling of anger, for instance, may arise in any situation where the belief or desire of the other is deemed to be in some sense devious. The results of that reaction may get quite nasty. They encompass many possibilities: from two children fighting over what to do to religious wars where the mutual anger is inflamed by the awareness of two sides having incompatible beliefs and preferences. Or take the disappointment and bitterness of a person who knows that some significant other doesn't share her beliefs and desires. Or, the despair of those whose beliefs don't find any resonance in the community they live in. The list here, as with Strawson's, is open-ended and I don't attempt to reveal all the possibilities.

---

<sup>37</sup> I will analyse further what it means to be alienated in the seventh chapter of this thesis.

<sup>38</sup> Naturally, there can be other factors.

<sup>39</sup> These claims about reactive attitudes can even be taken as *a priori* in a very weak sense, in that they characterize certain practical possibilities that belong to the everyday folk psychological practice the existence of which can't be directly falsified by empirical research. For a similar idea, see Heal 1998.

<sup>40</sup> For some suggestions concerning reactive attitudes towards beliefs, see Zimmerman (2007), and Schwitzgebel (2010). Both stress the rational responses of criticism and approval.

Reactive attitudes aren't limited to negative ones, of course, although the latter are somewhat more salient. Some desires and beliefs are more commendable than others and having them can also induce approval on the attributer's part. It is probably quite common that when someone discovers that another person shares a belief with her, she also finds that person to be more congenial and feels positive emotions towards her (Lydon et al. 1988). The same applies to shared desires the awareness of which increases the probability of achieving common goals, so that people should appreciate the similarity in attitudes. It is also possible in principle that people sometimes feel positive affect towards those whose attitudes are different from theirs but the reverse tendency is presumably still more prevalent. Similarities and harmony between persons are usually more attractive to both parties.

Although largely speculative, the philosophical point that can be picked out from these examples is that the reactions that the mental state ascription induces can be focused on the attributed attitudes themselves and not their behavioural or psychological consequences. What additional variables come into play is an open (and empirical) question but doesn't inhibit us from making a general claim. The affective importance of others' mental states as such provides an additional consideration to doubt the claim that prediction is the basic fruit of folk psychology.

But why do people have reactive attitudes to mental states in themselves?<sup>41</sup> I am here not seeking a sub-personal explanation, at least not at this point, because from the pragmatic perspective, a more interesting question is why the beliefs and desires of other people should matter to us in producing such reactive attitudes and how they relate to our practical interests more generally. One possible response would still be to maintain that we should care about them because of their *potential* behavioural consequences. Although sometimes they may not produce them it is reasonable for us to be concerned about them simply because they are possible. It's better to be safe than sorry and caring about the beliefs and desires seemingly non-instrumentally is just the most simple and efficient strategy to avoid being unpleasantly surprised by other's actions. In this way it is perhaps possible to conjoin the fact that the reactive attitudes are caused by the mental state attribution itself and the claim that one is still in a more fundamental sense interested in the possible behavioural outcomes of those states.

I don't think that such explanation is satisfactory. After all, I have already argued in the second chapter that the predictive potential that folk psychology offers is rather meagre and doesn't really distinguish mentalistic vocabulary from other kinds of conceptual tools. To claim that the main rationale for the mental state ascriptions is the making of predictions is to make an empty gesture in order to somehow stick to one's guns. Also, since I am here

---

<sup>41</sup> "In themselves" doesn't mean that mental states aren't attributed in a wider context of social circumstances and personal characteristics. It simply refers to the fact that our emotional reactions don't strictly depend on behavioural consequences of the mental states.

interested in an explanation that only attempts to put the mental state attribution in a context of our consciously acknowledged values and concerns; much in line with what I assume to be Strawson's approach (Strawson 1974: 24). Explaining the importance of our reactive attitudes in such an indirect way overlooks their most immediate consequences for us. What are they?

Keep in mind that my focus has been on incompatible, not compatible, desires and beliefs, and mostly on the negative reactive attitudes towards them. My answer to the previous question is mostly confined to such negative attitudes such as resentment and anger. Positive reactive attitudes to compatible mental states might require a different explanation. This shouldn't be seen as a flaw in my account since it isn't clear why different kinds of reactive attitudes should be explained in similar ways. So what explains the negative attitudes? I propose that the beliefs and desires of others – at least when they are different from the ones that the attributer has – function as a *challenge* to the attributer. For instance, the challenge posed by the divergent desire of another person isn't confined to the fact that this desire may cause actions the attributer may not like. In fact, I've argued against the view that the possible behavioural consequences are the main thing to care about. Merely the awareness that the other person wants something different demands at least an implicit response. The divergent desire also expresses the uncomfortable truth that the state of affairs that the subject cares about may be deemed unworthy by others because of their divergent desires and that is what gives the attributer a reason to be concerned; be it a bad reason or not. As for incompatible beliefs, the explanation is similar: being aware that another person has a different belief about some subject matter constitutes a challenge to the attributer because the latter needs to ask which of them is right in her belief. One needs to take into account that someone believes something different and that can be quite disturbing.<sup>42</sup>

In sum, both divergent beliefs and divergent desires function as challenges to the attributer of attitudes because they make her aware of the particularity of her cognitive and conative perspectives. This may seem to some to be a rather shallow explanation because it doesn't say anything about the causal mechanisms that bring about reactive attitudes in the face of others' mental states. Nevertheless, there is still some value in bringing out how the reactive attitudes to the mental state attributions are related to the wider concerns of the attributer – to know what is true and what is satisfying. The awareness of other perspectives calls that into doubt and can bring about negative affective reactions, especially if the attributee's perspective somehow dominates the attributer's own.

---

<sup>42</sup> This kind of depiction feeds quite smoothly into the debate in epistemology over conciliationism. According to that view, one should decrease radically her confidence in her belief if her epistemic peer disagrees with her (Christensen 2009). The clear difference here is that I don't intend to prescribe anything when saying that the awareness of a divergent belief can shake one's confidence and elicit negative reaction. Nor I distinguish between epistemic peers and other folk.

As I said, the reactive attitudes to compatible beliefs and desires might demand a different explanation. In this case, the mental states of the attributee function as an additional support for the attributer's own perspective, which tends to bring about positive affective reactions. Such an explanation wouldn't even be completely different from the one given to the reactive attitudes towards incompatible attitudes. While the latter challenge the attributer's own perspective the compatible attitudes confirm it, and challenge and confirmation can be seen as two opposites on a continuum. But I won't delve into that here. The important general idea here is that the common sense conception of mental states can be quite affect-laden and this is because the beliefs and desires of other people directly bear upon one's own theoretical and practical perspectives.

### **3.4 Manipulating and negotiating attitudes**

#### **3.4.1 Becoming active**

Thus far in the chapter we have only considered reactions to the mental states of others which are passive and don't require any deliberation or active thinking and planning on the attributer's part. In the case of reactive attitudes I mostly concentrated on the passive sentiment of *ressentiment*, which arose when the attributer wasn't able to do anything about another's belief or desire. Of course, situations where the other person can be at least to some extent affected by the attributer are probably more common. So what can the attributer do in a situation of conflict<sup>43</sup> between her mental states and those of others, besides simply having a reactive attitude or conforming to the other's perspective? What are the relevant possibilities here?<sup>44</sup>

To analyse those cases in which we are aware of the other person having a belief or a desire with which we are unhappy, it is useful to consider these possibilities in the form of a decision tree, starting from the most generally defined opposite reactions one might have and moving downwards to more specific possibilities. On every level, one of the reactions puts an end to the interaction game, while the other carries forward the dual structure. So, starting from above, the most general choice is simply whether to ignore the unattractive attitude or do something about it. Consciously ignoring such an attitude is a perfectly feasible option but it also puts a quick end to the flow of possibilities. This doesn't mean that the aforementioned passive reactions like automatic adoption or a reactive attitude couldn't still be available but these don't require any active deliberation on the part of the subject anyway. So the possibility of ignoring others' beliefs or desires is one way to react, and I am not going to speculate when it is cognitively demanding and when it is not.

---

<sup>43</sup> By conflict I mean that the subject considers the belief or desire of another person to be reprehensible in some way.

<sup>44</sup> Although conflicting beliefs and desires are in focus, I don't deny that situations when there is harmony between attitudes do not have their own practical possibilities.

The opposite of ignoring is doing something about the other's attitude, and this can be realized in two ways. The first of them, conforming to the attitude in question, reflects the automatic adoption already analysed above. The only difference here is that the subject *chooses* to take over the other's attitude, while considering, for instance, the fact that the other is in a position of power or that she has good reasons for holding that attitude. The weighing of reasons wasn't possible on the passive level. But the other option besides conforming which additionally evolves into two separate paths is what I call "alteration": an active attempt to change the attitude of another person. The question is now how to mould the desires and beliefs of others, and I distinguish two possible answers – manipulation or negotiation<sup>45</sup>. These are not mutually exclusive options – negotiation can include a noticeable amount of manipulation – but they are still analytically separable.

Manipulation, as I define it for the present context, is a way of leading another person into thinking or wanting something while concealing one's own role in producing that change or, if not concealing, at least not allowing the other to challenge this role and cast doubt upon it. This may not be an orthodox definition of manipulation, if there even is any<sup>46</sup>, but it should help us make the intended distinction between manipulation and negotiation. Negotiation should be possible only when the other is able to respond to the attempts of the attributer to change her beliefs and desires. Manipulation and negotiation should express two general ways of moulding the desires and beliefs of other people. In case of the first, only one person is the active participant, in the second, both are. Both of those two strategies can elicit reactive attitudes in others and may result in a rather automatic adoption of attitudes, as described above. Unlike the previously analysed responses, both manipulation and negotiation seem to require active deliberation on the part of the subject. And finally, the difference between these two strategies stems from the fact that negotiation also requires the other person to be an active participant in the process. The manipulation is also a two-way game in the sense that the subject has to calculate the possible responses of the other person and his possible moves, depending on those responses. But the fact that the manipulative moves of the subject are dependent on the other person doesn't mean that the latter can challenge those moves, which is necessary for negotiation to emerge.<sup>47</sup> In what follows, I will briefly analyse each of these strategies.

---

<sup>45</sup> Or "persuasion", as Rudinow (1978: 338) puts it. He also adds coercion as the third possibility. Since I am interested here in ways in which people try to influence beliefs and desires of each other, I leave coercion out of my account because it aims rather straightforwardly at influencing actions and overlooks the motivation of a person.

<sup>46</sup> There seem to be two uses of "manipulation" in everyday talk, one is morally neutral and the other is not. I prefer the first one but I admit that using this expression may bring about unwanted connotations.

<sup>47</sup> The distinction between negotiation and manipulation is basically the age-old opposition between authority and mere power, reason and persuasion, argument and force (Williams 2002: 9) which are often hard to distinguish.

### 3.4.2 Manipulation of attitudes

Probably the most obvious example of manipulation is deception. The growing ability to deceive conspecifics is seen by many to be one of the underlying causes of the evolution of social intelligence and intelligence in general in primates (Byrne & Whiten 1988). The ability to deceive isn't even confined to those organisms which have the ability to ascribe mental states. This is demonstrated by the simple phenomenon of mimicry, for instance. In the present context, the evolutionary role of deception isn't the most pressing issue. The existence of intentional deception, the aim of which is to induce a false belief in the other person, is surely a real phenomenon in the social world, and has a part to play in social situations. Deception can proceed through different paths – by giving the other false information, by holding back relevant information, but also by offering true information which is formulated so that it is possible to make some false inferences from that.

Is it also possible to deceive someone into wanting something? There are no false desires, only inappropriate and failed ones. So deception doesn't seem to apply to them. Yet one can still talk about it when a person induces in another a desire that she originally didn't have and does that without giving her an opportunity to consider the advantages and disadvantages of having that desire. Desires can still be responsive to reasons, at least to some extent. But it also seems that deceiving someone into wanting something requires the prior manipulation of her belief about what is worthwhile to desire and what is not. In that sense, the deception is primarily about beliefs.

Deception is also a rather specific way to influence the beliefs of others because it only aims at creating false beliefs. Manipulation, as I defined it, isn't confined to that because it involves any attempt to influence other's attitudes without letting her actively contribute to the process. So it is perfectly possible to manipulate someone into believing something true, as long as the other person is unaware of the process or unable to control it. The same applies to desires – people can be made to want something that is actually useful to them, the process still being a form of manipulation. Take the adoption-by-attribution model, for instance, that was considered before, according to which people can adopt the beliefs and desires that they have attributed to their social superiors. Even this can be taken as manipulation by the latter if they intentionally exploit their status in inducing attitudes in others, whether these attitudes are true and useful or not.<sup>48</sup> More generally, if a true belief or useful desire is formed because of the action of another person without the subject being able of even minimally reflecting on that, this can be taken as manipulation.

Joel Rudinow has also argued that manipulation isn't limited to deceptive moves. For him, manipulation aims at changing the motivational profile of a person by deceiving her *or* using her weaknesses (Rudinow 1978: 346). The second disjunct is about creating in another person an incentive that she can't

---

<sup>48</sup> This isn't meant to deny that adoption-by-attribution is mainly an automatic and nonintentional process.

resist. In a way this kind of disjunctive account captures the distinction between manipulating another person's beliefs and desires, respectively. Manipulation of beliefs would be a form of deception and the manipulation of desires would be a way of exploiting the other's weaknesses. This is still a bit too restrictive because Rudinow seems to confine at least the manipulation of beliefs to the formation of false beliefs while my account tries to avoid this implication. But the general spirit of his approach is pretty similar to the one considered here.

But is my definition of manipulation – a definition that is not limited to deception and includes any kind of one-sided social influence – satisfactory? Is it an unhappy consequence of the definition that only reflectively acquired beliefs and desires can be seen as not being outcomes of manipulation? After all, there are probably many beliefs and desires that originate from others but are not formed with any explicit awareness or ability to control them. Most of the attitudes of children are probably acquired from others unreflectively. Does that mean that these attitudes are also the objects of manipulation? Or what about the case of rhetorical persuasion which relies not only on arguments but also on particular qualities of speakers and listeners?<sup>49</sup> Since I am not aiming at drawing a moral distinction here, I don't think it's really a problem. The pedagogy of children does have a strong manipulative element, in the sense of manipulation considered here, because it often avoids any active contribution from the pupil, at least in the earliest years of a child's life. I suspend judgement on the question whether this kind of manipulation can be avoided or not, or even whether it is good or not. As for rhetorical persuasion, one can also say that it is a form of manipulating others into believing and wanting something. In public discussions manipulative elements are intermingled with negotiation, and the distinction presented here is, in a way, an ideal one.

Both negotiation and manipulation can be ways of being affected by mental state attributions. The main difference between the two, as I've already noted, is that in negotiation the interpreted person plays an active part in responding to the interpretation. Still, what significance does the distinction have for the folk psychological interpretation and affectability? The idea of affectability was meant to bring out how others' beliefs and desires can bear upon us and not in the way that physical facts do. The awareness of them can affect us in specific ways – through reactive attitudes, adoption of others' beliefs and desires, or our own actions – without necessarily any predictive inferences concerning the direct behavioural consequences of those attitudes.<sup>50</sup> In manipulation, our actions in relation to others' attitudes may be somewhat more similar to our dealings with merely physical facts or objects than in negotiation. After all, manipulation is usually associated with using mere force and rather than reason

---

<sup>49</sup> For Aristotle, for instance, there are three sources of persuasiveness in rhetoric, the first comes from the character of the speaker, the second from the emotional state of the hearer, and the third from the argument (Rapp 2010). According to my account, only the last one would be a form of pure non-manipulative strategy.

<sup>50</sup> I am not denying that we make predictive inferences concerning the possible effects of our actions on the mental states and behaviour of others.



or argument. The manipulator seems to treat the other as an object of purely causal treatment.

Although this depiction isn't completely off the mark, there is still an element in the manipulative behaviour that is important to point out. If the objects of manipulation are beliefs and desires, the sheer-force metaphor isn't appropriate because it doesn't seem to be possible to simply force people to believe or want something, even if one is in a position of power.<sup>51</sup> This is largely an intuitive claim and may be disproved in the future when the neurosciences evolve to the point at which its means will be available for many. But at least for now, the manipulation of others' beliefs and desires is still a difficult skill that requires the ability to exploit the complex web of social relations.

### 3.4.3 Negotiating attitudes

As I've postulated, what distinguishes negotiation from manipulation is that the other person plays an active part in the interpretation process by being able to challenge the attributer's role in inducing in her the belief or desire in question. Negotiation is reciprocal in that both sides play their part in construing the folk psychological interpretation. So it is not necessary that the attributer has a pre-given understanding of what attitudes the other has and a pre-given plan that specifies how to change them. Such an anticipatory understanding is in any case unstable because what mental states the other has may become clear only at the end of the negotiation. Furthermore there may be no specific initial plan that concerns how to change the other's attitudes because the attributer's understanding of her communicative intentions may be ambiguous and only become clear during the process of negotiation (Gibbs 2001: 109). The attributer's own mental profile in general may change during the process and the participants may come to have attitudes that they didn't know that they have had before (Ratcliffe 2006: 39). That is the most important reason why negotiation is a specific way of being affected by folk psychological interpretations because the interpretation of the other can also influence the self-interpretation of the attributer and she might come to change her own self-understanding. In the case of adoption-by-attribution something similar is going on because the attribution can also result in an attitudinal change but that process is nonreflective and nonreciprocal. If the inclination of the attributer to adopt the attributed attitudes is exploited by the other person who is in the position of power, manipulation emerges. But in negotiation, both participants have some control over the process, which is also more dynamic.

In the case of beliefs, consider, for example, someone who attempts to change a belief which she has attributed to another person by presenting some considerations against it. Since the other is in a position to respond to that attempt, the process of negotiation may go in different directions: if the other

---

<sup>51</sup> Remember also the remarks in the previous subchapter about the instability in the power over attitudes because of the *ressentiment*.

doesn't agree with that attribution she can argue that the interpretation itself is mistaken, or, if she agrees, she can take into account the arguments of the subject against that belief. The outcome may be that the original attributer herself adopts the belief in question if the other manages to convince him in the second way. Alternatively, the attributer may simply drop the attribution if the other is able to make the case that she actually doesn't have that belief. This may seem obvious and even trivial but it elucidates nicely the way in which the process of folk psychological interpretation – when reciprocal – is something quite different from explaining and predicting the behavior of physical objects. As for desires, negotiation may seem to proceed differently, if at all, because it isn't clear whether desires can be responsive to arguments. If negotiation is about challenging someone's attitude, having a desire may seem to be immune to that kind of challenge. Also, can one speak about justification in the case of desires? If one is a Humean about reasons and takes them to be derivable from desires in some way, then desires are a bedrock of justification in practical reasoning and asking for reasons for having a desire is an inappropriate manoeuvre.

The first thing to note is that even if desires are not responsive to reasons, it is still possible to maintain that in negotiation the *attribution* of a desire can be challenged and interpretation is still reciprocal in that sense. Even if there are no reasons for having a desire, arguments for or against ascribing a desire to someone still have their place in social interaction. But, although I wouldn't delve too deep into theories of practical reasoning, I still claim that the idea of desires being responsive to reasons has some plausibility when we consider everyday conversations between reflective individuals. It seems to be entirely appropriate to ask such a person if she should have the desire she has or whether the desire in question is actually useful for her. If she admits that the desire she has isn't actually beneficial by any means, she may be able to give up that desire, at least in the long run and in the absence of addiction or some other abnormality (see Moran 2001: 115). If that is true then the difference between the reason-responsiveness of beliefs and desires may be a mere temporal matter and desires can be negotiated in the same way.

The account of negotiation presented here hardly amounts to a completely original claim because traditional approaches to folk psychology (theory theory and simulation theory) have already been criticized by those authors who I will call interactionists<sup>52</sup> and who see the social interaction as a crucial or even constitutive aspect of folk psychological understanding (Gallagher 2006, Hutto 2004, Ratcliffe 2007). I also see negotiation as a counterexample to the detached prediction-explanation model of folk psychology because in negotiation the interpretation proceeds through reciprocal interaction and the attitudes can often become aims in themselves and their possible behavioural consequences may have only secondary importance. Nonetheless, I am not making any claim here that negotiation (or interaction) is a necessary

---

<sup>52</sup> This may not be the best label for them but I can't think of any better.

component of mental state attribution because there are other ways of being affected by mental state attribution, which I have introduced before. The aforementioned authors tend to generally overlook manipulative forms of social cognition.

Still, the affectability through negotiation does seem to assume the mutual responsiveness of the interpreter and the person being interpreted. When speaking about others in terms of beliefs and desires we can position ourselves to those attitudes before knowing their future behaviour so that we can act in response to those attitudes. It is natural to assume further that we can treat the other who we interpret also as the one who is interpreting us. Two interpreters have, at least to some extent, different points of view so that taking into account the other's perspective cannot proceed solely from a projection of one's own perspective on the situation. If the interpretations are mutual then the other can also help or thwart our interpretation of them and also our interpretation of ourselves. If I attribute a belief or desire to someone she can challenge my interpretation, and vice versa. If we are on to something with the idea of affectability, and the folk psychological attribution's relevance comes before its predictive function, then the possibility of mutual interpretation doesn't depend on it either.

To give an additional example of an account of interpretation through negotiation, James Bohman (2000) connects the mutual interpretability with the second person perspective which consists of taking an appropriate performative attitude and which differs from both first- and third-person perspective in that it depicts the person being interpreted as an active partner in a dialogue who is also able to respond to my interpretations. He sees it as a distinctive stance toward other persons when we want to understand them and he even postulates it as a methodological standpoint that the social sciences should take.

It is not *I* or *we* that must be satisfied; the interpretive situation is one in which an interpreter is interpreting someone else, who in turn can offer an interpretation of his own actions and the interpretation of the other's interpretations. The *I* who is an interpreter is a *you* as a participant in communication, and thus no account of the hermeneutic circle can do without the second person. (Bohman 2000: 223)

I am not in a position here to judge the role of negotiation in the social sciences but Bohman's account of mutual interpretation is relevant in the present context because it exemplifies the reciprocal element in negotiations and how they affect the subject by (re)considering also her own self-interpretations. In conclusion, knowledge of others' beliefs and desires gives us a possibility of negotiating their views of the common world or to from that sort of negotiation. Both constitute a response, even if the latter does so only in a negative way.

### 3.5 Taking stock

The possibilities presented here were meant to provide an elaboration of the affectability that comes from the folk psychological attributions, which is hard to account for in terms of a simplistic explanation-prediction model. The reason why the latter faces trouble is that the mental states that we ascribe matter to us not solely for satisfying explanatory or predictive concerns. Instead, the possibilities considered here – adoption-by-attribution, reactive attitudes, manipulation and negotiation – revolve around people having those attitudes, without necessarily aiming at the behavioural consequences of those mental states.

These considerations don't tell us much about *the* function of folk psychology but still reveal its substantial consequences for human social life. They don't tell us about the function because both what folk psychology was selected for, and why it emerged in the first place, are still open questions. I can in any case claim that the meaningfulness of that question doesn't affect the points presented here. Even if folk psychology initially emerged because of the behavioural consequences that were inferred from attributions of mental states, now the attitudes themselves can be non-instrumental objects of concern (see Woodward 2009). How this can be was exactly the question that I have been trying to answer here. Now it is time to draw these different threads together.

When the subject attributes a belief or desire to another person, it is quite safe to assume that the different possible reactions depend on the particular relationship between those agents.<sup>53</sup> In the end, it is the job for social psychology to clarify the most crucial variables that play a part in the function of affectability. Still, I've mostly stressed the role of social power between the attributer and the attributee, although trustworthiness in the case of belief-attribution and attractiveness in case of both mental states may also be important. Then again, attractiveness can also be taken as a form of social power since the latter is a rather wide category. When there is an inequality between the interpreter and the one being interpreted, so that the first dominates the second, the result may be the automatic adoption of the attributed mental state by the subject and/or *ressentiment* towards the other person because of the attributer's inability to influence the other's attitude. If the inequality in power goes the other way, the subject is in the position to somehow manipulate the other into believing or wanting in a manner to her liking. And finally, when the power to influence the other is relatively equal between two persons, the most natural response is the attempt to negotiate the attribution if one party is dissatisfied with it. On the other hand, if the negotiation happens in the situation of inequality, the person in power can push it towards manipulation again.<sup>54</sup>

---

<sup>53</sup> What is lacking in my account, is attention to group-level attributions, but I do think that the most paradigmatic mental state attributions take place in two-person interactions.

<sup>54</sup> To anticipate a possible question, I should note that the reason why I have mostly focused on the possible reactions to *divergent* attitudes comes from the fact that if the attitudes of persons agree with each other the participants find it less necessary to adopt a metarepresentational stance and they can remain at the first-order mental states, directed at the world.

One could raise the concern that I haven't really defined what social power means here and I need to respond because this concern is perfectly reasonable. In the most general sense, social power is the capacity of a person to influence the behaviour and mental states of other persons, whether or not they agree to this influence. Stated that way, the definition is certainly very vague but perhaps satisfactory for present purposes. It is still possible to make some specifications. For example, social power shouldn't be conceived merely as a power to do something (Hobbes 1996 [1641]: 58), but as a power over someone (Dahl 1957: 202)<sup>55</sup>, being always interpersonal. Also, power is a dispositional concept (Lukes 2005: 69), a capacity that can be actualized only in specific conditions

An interesting question is: to what extent the social power of a person is also constituted by the folk psychological attributions of others? For instance, if someone feels *ressentiment* towards another person and is inclined to adopt her mental state because of that, contrary to her own preferences, the social power of the other not only influences the way in which the subject is affected but it is also partially constituted by the affectability by which the subject is moved by the other. The fact that the others are inclined to adopt the attitudes of a more powerful person makes the latter even more socially dominant. Power can thus be gained by controlling the interpretation of one's mental states by others if it brings about appropriate reactive attitudes and submission to the mental profile of the subject. Because of that, the affectability that comes from folk psychological attributions in unequal situations can enforce already existing power relations and even create a new kind of social power; the one that is rooted in persons being aware that the other (the dominant one) has certain mental states which they cannot influence.<sup>56</sup>

Following this line of thought, power can also be gained through the ability to create second-order desires in others that have one's own beliefs and desires as objects. If others want the subject to have certain attitudes but are unable to influence them, they are stuck in a situation where they don't really have many options besides either conforming to those attitudes themselves (adoption-by-attribution) – and thus basically giving up the second-order desire – or remain in impotent resentment (*ressentiment*). In both cases, the subject retains her dominant position because the others depend on her attitudes (or, more precisely, on their interpretation of her attitudes). The subject who is able to generate such second-order desires in others knows that they are affected by their own attributions and can take advantage of this knowledge.

Power relations influence the way in which persons are affected by their folk psychological attributions, and *vice versa*, but do they also affect the attribution

---

<sup>55</sup> In Dahl's definition, A having power over B implies a counterfactual so that B wouldn't do certain things if A hadn't any power over him (see Lukes 2005: 44). I tend to think that this is too restrictive. Although it may be epistemically more difficult to understand the situation where this condition isn't met, it is still possible to talk about A having power over B, as long as she is able to influence her B in some relevant way.

<sup>56</sup> How the ability to control others' attitudes the highest form of social power is also stressed by Lukes (2005: 27).

process itself? Do they play a part in determining how people attribute mental states to others? This is a further question, one that I won't answer here. But the take-home message at this point is that the affectability that folk psychology provides is closely entangled with the questions of social power and influence and its effects can be quite different depending on the specific relationship that the subjects are involved in. Through folk psychological attributions people consolidate and disentangle their friendships and animosities, which feed back into the effects that those attributions have. All this can centre on mental states themselves, not only their behavioural consequences.

In this chapter I elaborated on the idea of affectability and hopefully succeeded in bringing out the diversity of reactions to folk psychological attributions. I presented them in an order that began with the more automatic ones and I worked towards those which include reflected negotiations. However, in all cases, the social relationship between the attributer and the attributee plays its part in influencing the intersubjective consequences of the attribution process. In what follows, I hope to show that this aspect of folk psychology is also relevant to the way that we construe the application conditions of folk psychological concepts.

## 4. FROM AFFECTABILITY TO THE COMMUNICATIVE CONCEPTION

### 4.1 Introduction

In previous chapters, I analysed what I called the affectability of mental state attributions which revealed itself in the responses of agreement/disagreement and endorsement/disapproval to the attributed attitudes and the further reactions that derive from them. Although the analysis stayed at a descriptive level, I also argued that such engaged responses are indispensable for our folk psychological practice. That is to say, such responses are always open to the attributers of attitudes. Affectability is something that can't be reduced to other benefits that mental states attributions provide such as prediction, explanation or mind-shaping. Nevertheless, one may ask whether this fact implies anything interesting about the concepts of belief and desire or even about the beliefs and desires themselves.

In this chapter I am going to seek a construal of mental state attribution that makes intelligible why the engaged responses to the ascribed attitudes are appropriate. This sharpens the understanding of the concepts of belief and desire and it will raise the question what their application conditions must involve in order to explain the indispensability of the engaged responses. I will focus on affectability in the narrow sense, which includes the responses of agreement/disagreement and endorsement/disapproval. By "application conditions" I mean the conditions under which the attributer counts as having applied the concept in question.<sup>57</sup> In the case of beliefs and desires, then, the question is about the conditions under which the concepts of belief and desire are applicable to persons in a way that can deliver the benefits of affectability. I label this way of analysing the application conditions "explication". By using that term I don't intend to commit myself to Carnap's (1950) project of conceptual explication with which the present account might be associated. I don't want to engage in historical exegesis or to ascertain whether there are substantial similarities or differences between our accounts. Because hopefully this doesn't matter for the present project. A more directly comparable account is Justin Fisher's (2006) project of pragmatic conceptual analysis which operates along broadly similar lines and to which my own account owes a lot. In his case the explication also starts from a consideration of the main benefits of a concept and then "reverse engineers" it to see what its application has to involve in order to provide those benefits. As far as I'm aware, though, he himself hasn't applied his approach to the concepts of belief and desire. What matters here is the possibility of inferring something substantial from the benefits of a certain way of concept-use to the application conditions of those concepts.

---

<sup>57</sup> In the present case one could also talk about the attribution conditions of beliefs and desires.

The pragmatic conceptual analysis (or conceptual explication) differs from the standard conceptual analysis by not requiring that the specific conditions of concept-application correspond to the intuitions of the concept-users. As long as those application conditions can explain how the concepts in question deliver the benefits that we are interested in, the explication can be satisfactory. I consider this to be an advantage of explication over the standard conceptual analysis because it isn't clear that intuitions are the best guide to the meaning of our concepts. I won't argue for such skepticism towards intuitions in this chapter, though. Even if conceptual analysis were a trustworthy method for understanding the application conditions of our concepts, the present exercise would still be worthwhile in its own right. As long as we are interested in understanding how our concepts can bear the fruits they do, explication deserves some attention. It helps us to see how our concepts can benefit us even under possibly unintuitive application conditions. The explication doesn't aim at capturing all the main patterns of use but at clarifying the conditions under which the use gives us the benefits we care about. It can thus have revisionary repercussions: certain commonly acknowledged features of the concept's use may be rejected if they are not relevant to the delivery of the benefits in question.

Still, the content of the application conditions should be epistemically accessible to ordinary ascribers. Only then would they be in a position to know when these conditions hold because the ascribers should be able to consciously put to use the results of the explication. Otherwise the whole endeavor wouldn't make much sense. The explication is meant to help the users of those concepts understand what they are doing and why it is worthwhile. There might also be a substantial psychological story about how the application of concepts works. However, since the workings of cognitive mechanisms themselves are not accessible to subjects, the application conditions shouldn't be expressed at that level. The hope is that after we have explained how the main benefits are delivered, there will be nothing more at the personal level to be said about the application conditions of the concepts in question.

I will proceed as follows. First, in section 3.2 I will consider and reject certain accounts of the application conditions of the concepts of belief and desire. This diagnosis should help us identify a better account that would explain the benefits related to affectability. In section 3.3 I will consider the possibility of understanding folk psychological attributions in terms of speech acts. Such an approach allows us to see how the application conditions of those concepts can be modeled on the conditions for uttering and comprehending assertions and requests. Then, in section 3.4, I will analyse Christopher Gauker's theory of propositional attitude ascriptions (the communicative conception) which is the best (and only) version of this kind of approach. Finally, in section 3.5 I will present an additional consideration in support of Gauker's account. I will defend the view that his communicative conception provides the necessary application conditions for the concepts of belief and desire for explaining affectability. Discussion of the shortcomings of the communicative conception will be left for the next chapter.



## 4.2 Initial proposals

When we now think about the proposed main benefits of the concepts of belief and desire, namely, the responses of agreement/disagreement and endorsement/disapproval, it might seem difficult to get the explication started. What could the proper application conditions for those concepts be if the use of those concepts is to deliver the effects in question? In order to fully articulate these responses as benefits of the concepts in question we should also bear in mind affectability in the wider sense. This includes the emotional responses as well as the negotiation and manipulation of others' attitudes. Does the possibility of such engaged responses put some constraints on how we should understand the application conditions of mental predicates?

In what follows, I go through some well-known theories of mentalistic concepts to see how well they are able to explain the indispensability of engaged responses. I don't intend to seek a deductive argument in support of one account against others. The aim is to make an inference to the best explanation which should satisfy three desiderata. First, it should make explicit why engaged responses are not merely contingent effects of mental state attributions, but need to be open to an ascriber whenever beliefs and desires are attributed (as argued in the second chapter of the thesis). Second, a satisfactory account of application conditions should also show why beliefs and desires merit *different* responses of agreement/disagreement and endorsement/disapproval, respectively. Beliefs and desires have distinctive response profiles and acceptable application conditions of mental predicates must respect that. Third, the proposed application conditions should be epistemically accessible to ordinary ascribers: there shouldn't be any need to defer to experts in order to know whether these conditions hold. This also means that the ascribers themselves should be able to understand why affectability is a proper response to beliefs and desires.

So let's consider certain theories of mental concepts and see whether the application conditions that can be read off from them can explain the indispensability of affectability. Let's take, for instance, behaviourist accounts that attempt to analyse the application conditions of mental predicates in terms of certain behavioural dispositions. On that view the concepts of belief and desire are applicable to the subject when she has the appropriate behavioural dispositions (Ryle 1949).<sup>58</sup> When we ask whether the behaviourist account can explain the indispensability of the engaged responses to the attributed attitudes, it isn't clear how she can do that. For the behaviourist, the ascription of attitudes only commits the ascriber to considering the pertinent dispositions and there is no reason to think that she should be able to agree or disagree with the attributed belief or to endorse or disapprove of the attributed desire. Her only forward-looking response is prediction of behaviour on the basis of the

---

<sup>58</sup> Although behaviourism has fallen out of favor, dispositionalist views which also analyse mental states in terms of dispositions, not necessarily only in terms behavioural ones, can still be found on the market (see Schwitzgebel 2002).

conditionals which specify the pertinent disposition. So if I am right and the engaged responses are indispensable for propositional attitude attribution, behaviourism has difficulties in accounting for that. If we take it as a theory about the application conditions of the concepts of mental states, it doesn't manage to function as an explication that properly registers affectability as the central benefit of the propositional attitude attributions.<sup>59</sup>

The problem with behaviourism seems to generalize to functionalist views about the concepts of mental states. Let's ignore the plethora of different views that go under the label of "functionalism." These accounts in general, as long as they even profess to say something about the application conditions of the concepts of belief and desire, construe them in terms of the causal relations between environmental inputs, inner states and behavioural outputs (see Lewis 1972). Yet beefing up the information base of the application conditions with reference to inner states doesn't seem to make functionalism more successful than behaviourism vis-à-vis their capacity to explicate the concepts of belief and desire. It still doesn't explain why the engaged responses should be indispensable to the attributers. Predictions and explanations seem to be the only necessary benefits that knowledge of the functional roles of mental states provides.

We are now in a position to diagnose the failure of behaviourism and functionalism and draw a more general conclusion from that. It seems that no account of the application conditions of the concepts of belief and desire that takes the use of those concepts to consist of merely identifying certain regularities or causal relations between inner and outer states of a person can account for the indispensability of engaged responses of agreement/disagreement and endorsement/disapproval and thus fail to satisfy the first desideratum of acceptable application-conditions. Any explication of the concepts of belief and desire which aims to explain such indispensability needs to introduce something extra in order to make explicit the connection between application conditions and engaged responses for an ordinary ascriber.

With that in mind, a rather obvious move would be to turn to representationalist views which take propositional attitudes to be relations to representations with propositional content (Fodor 1975; Cummins 1996). After all, such representations are semantically evaluable and engaged responses of agreement/disagreement and endorsement/disapproval consist of taking a stand on the content of another's attitudes. If the application conditions of the concepts of belief and desire make reference to the contents of those attitudes, the engaged responses seem to be an appropriate possible outcome of the mental state ascriptions. Since functionalism and behaviourism don't make such

---

<sup>59</sup> Note that I am not here considering the nature of attitudes, but only the respective concepts. Inferring something about the latter from the affectability of belief- and desire-attributions would be a too hasty step. It might be even true that behaviourism is correct about the nature of beliefs and desires but mistaken about the application conditions of the concepts.

reference in their application conditions<sup>60</sup> one could claim that they failed to explicate the concepts in question. On the other hand, if the attribution of beliefs and desires involves picking out the representations which are semantically evaluable then the engaged responses may seem to be understandable reactions to the ascribed attitudes. I am not saying that a representationalist is necessarily committed to claiming that the attributer of beliefs and desires has to identify the mental representation which the attributed attitude is supposed to stand in a relation with. She may treat the nature and the attribution of attitudes separately. I am only pointing out that if she *does* make such a claim she might have a chance of explaining the indispensability of engaged responses. Representationalism, then, seems to satisfy the first desideratum, namely, that engaged responses to propositional attitudes are indispensable. Semantic evaluability, after all, seems to be an essential characteristic of mental representations.

This appearance might be illusory, however. It isn't clear whether a mere fact of tokening a mental representation can call forth an engaged response or why does it even matter to a person who is aware of this representation. As I've argued in the previous chapter, reactive attitudes that we have towards others' beliefs and desires should be explained (on the personal level) by the fact that beliefs and desires function as confirmation or challenge to the ascribers' cognitive and conative perspective, respectively. Unless a mental representation is backed up by communicative force (being used to claim that something is or should be the case), it doesn't provide any reason to have an affect-laden reactive attitude towards it. This is only a phenomenological point, though, and won't probably convince a representationalist.

However, representationalism also faces difficulties with the second desideratum of acceptable application conditions, namely, to explain the difference between beliefs and desires in such a way that would explain why the engaged responses to those two types of attitudes are different. Attributing propositional content doesn't yet constitute attributing beliefs or desires. One way for a representationalist to distinguish between the two is to claim that although belief that *p* and desire that *p* have the same contents they are distinct because their contents enter into different functional roles. But in that case the distinction is based on functionalist considerations and we saw that the indispensability of engaged responses couldn't be explicated on their basis.

A different proposal would be to argue that beliefs and desires are actually relations to different types of content. For instance, one could distinguish between indicative contents which have truth-conditions and imperative contents which have satisfaction-conditions. Indicative contents specify what the world is like and the imperative contents specify what it should be like. Let's suppose that beliefs are relations to indicative contents and desires are relations to imperative contents. The move from the difference between the two kinds of contents to the difference in engaged responses could be made as

---

<sup>60</sup> Since functionalism is a rather general view about the mind, the idea that beliefs and desires involve relations to representations isn't strictly speaking inconsistent with it.

follows. The responses to the attributed beliefs, agreement and disagreement, can be taken as acts of affirming and denying what is the case. The responses to desires, endorsement and disapproval, on the other hand, evaluate what should be the case. If this is correct the difference in engaged responses seems to map nicely onto the difference between the two types of contents. If this suggestion works, the different engaged responses (agreement/disagreement and endorsement/disapproval) seem to latch onto two different types of content, thus satisfying the *desideratum* of the explication. In a nutshell: the response of agreement/disagreement is appropriate in the case of an attributed belief because the ascriber has assigned indicative content to it; the response of endorsement/disapproval is appropriate in the case of attributed desires because the ascriber has assigned imperative content to it.

Imperative contents are contentious. They can be found in Millikan (1984: 97) and in the work of some authors have tried to understand the content of pains in these terms (see Klein 2007; Martinez 2011, for a critique, see Cutter & Tye 2011). In the present context, it is worth pointing out this controversy because, assuming that the imperative theory of pain is successful, one can ask what distinguishes desires from pains. They seem to be distinct. It doesn't make sense to endorse or disapprove of pains on the basis of their contents. One natural proposal would be to articulate the distinction in terms of different functional roles. But then we would fall back into the earlier problem of explicating the engaged responses on the basis of functional roles – a problem which seems to lack a satisfactory answer. Furthermore, the threat of indistinguishability may remain if we reject the imperative theory of pain because there are other mental states that can be characterised as having imperative contents such as intentions, decisions or tryings. If we did need to turn to functionalist considerations in order to distinguish these mental states from desires then once again it wouldn't be clear how we could explicate the concept of desire merely in terms of imperative contents. The appeal to such contents doesn't seem to fully explain why the attribution of desire delivers the benefits it does, even if it involves assigning an imperative content.

Representationalism faces an additional problem. It doesn't seem to be able to satisfy the third desideratum, according to which the proposed application conditions have to be accessible to ordinary ascribers. 'Mental representation' is a technical notion and not something that belongs to the conceptual repertoire of the folk. In any case, ordinary ascribers are probably unable to recognize when a particular mental representation is tokened. This means that representationalist application conditions would be uninformative to them and wouldn't help to understand when it is appropriate to ascribe beliefs and desires. This difficulty that representationalism has in explicating the concepts in question shouldn't be surprising, though. Construing the application conditions for ordinary ascribers hasn't really been the central concern of its proponents.

A disgruntled voice might now object that I am simply too demanding. None of the proposed accounts we have considered (behaviourism, functionalism or representationalism) have explained the indispensability of engaged responses.

Still, none of them has been inconsistent with the fact of these responses either. Maybe the possibility of differential engaged responses need not be built into the attribution conditions. The fact that these responses are always open to us may simply be a psychological or social fact about us as human beings. Since evaluating what others think or prefer matters to people, this might suffice for explaining why we tend to have the responses in question. But this, the objection might go, doesn't require construing the application conditions of the concepts of belief and desire in the manner that explains the indispensability of agreement/disagreement and endorsement/disapproval. It is possible to cash out the application conditions in functionalist terms, for instance, and still maintain that the engaged responses are crucial, both psychologically and socially.

This objection has its force because the envisaged scenario seems to be possible. One cannot deny that merely functionalist application conditions, for instance, can deliver the explanatory and predictive fruits and *don't exclude* the possibility of engaged responses. It can deliver them as long as these application conditions can specify the contents of the attitudes and the difference between them, either on the content- or the attitude-level. One just needs to postulate in addition the (psychological) disposition to have two kinds of engaged response to those two kinds of attitude. But merely pointing out this possibility doesn't show that we shouldn't seek an explication that has the indispensability of engaged responses built into the application conditions. If it's true that affectability is indispensable for folk psychology, as argued in the second chapter, it is an explication we should strive after. The aim of the present chapter is to argue that there *is* an explication that explains why the engaged responses are always open to us (i.e. their indispensability), something that the other accounts can't fully achieve.

### 4.3 From engaged responses to speech acts

In order to reach the desired explication I need to take a detour. Remember that in the previous section I discussed imperative contents, as opposed to indicative contents, in order to help us draw the distinction between beliefs and desires and explain the difference in the appropriate engaged responses. It is noticeable that imperative contents are modeled on imperative utterances, and indicative contents on expressions in the indicative mood. Let's take a representation with an imperative content  $p$  which thus has a certain state of affairs or event-type as its satisfaction-condition. It is natural to express this representation in terms of the imperative utterance " $p!$ ", or "let it be the case that  $p!$ ".<sup>61</sup> Analogously, the expressions in the indicative mood should be applicable to indicative contents. A theorist may stress that the possibility of expressing the content in linguistic

---

<sup>61</sup> Perhaps instead of imperative, we should speak about optative utterances because imperatives are tied second-personal interactions, while the optative can take a third-personal form, too. Nevertheless, I stick with the term "imperative" and simply stipulate that it can be also be in the third person.

form doesn't imply that the representation itself has linguistic structure. Nonetheless, we at least model the representations in terms of linguistic utterances. It is actually very common to model mental states in such a way because the content of mental states is usually understood in terms of truth-conditions and reference. Linguistic utterances are exactly the paradigmatic entities that have such properties.

Notice that when we distinguish between indicative and imperative moods in language we can have the engaged responses, which were supposed to characterize beliefs and desires, also to assertions and requests. After all, these speech acts have satisfaction conditions, closely parallel to beliefs and desires, and we can frustrate them, satisfy them, negotiate with those who have performed them, etc. To put it more precisely, in the case of another person making an assertion that *p* my initial response to it is agreement or disagreement. The pertinent responses that might follow are basically the same as those that follow from the attribution of belief: an expression of solidarity or an attempt to persuade the other of being mistaken. The attribution of desires, on the other hand, mirrors the understanding of requests: if another person has requested that *p* I can endorse or disapprove it. If I endorse it I can try to fulfil the request. If I disapprove of the request I may try to act against it or simply ignore it. The pertinent responses to others' mental states mirror the responses to the respective speech acts: endorsement or disapproval, agreement or disagreement.<sup>62</sup>

Remember the difficulty that the representationalist had in distinguishing between the proper response to beliefs and the proper response to desires in providing their application conditions. This difficulty was tied to the fact that the possible differences in causal-functional roles didn't seem to provide any reason to think that this difference necessitates the difference in engaged responses. If we now consider the possibility of modeling the application conditions of beliefs and desires on assertions and requests then the different engaged responses map nicely onto those two kinds of speech act. To elaborate, the (sincere) assertion that *p* can be taken as the asserter's expression of her commitment to the truth of *p* and the (sincere) request that *p* expresses the commitment to the claim that *p* should be the case. This makes it understandable why the response of agreeing or disagreeing whether *p* is the proper response in the first case and endorsing or disapproving of the prospect of *p* is proper in the second. We have the same pattern of responses in the case of the respective attitudes.

---

<sup>62</sup> The response profiles to attitudes and speech acts are also similar with regard to them both functioning as a challenge or confirmation to the attributer's conative perspective in the case of desires and requests, and a challenge or confirmation to her cognitive perspective in the case of beliefs and assertions. They function as a challenge in the case of incompatible and as a confirmation in the case of compatible attitudes and speech acts. In the previous chapter I already brought out how attributed attitudes function as a challenge to or a confirmation of an attributer's perspective and this explains why they matter to her. Now the further idea is simply that the significance of speech acts for a person who comprehends them is similar to that of attributed attitudes.

Remember also that the appeal to the distinction between imperative and indicative contents was perhaps the most promising move for the representationalist to make. Yet it had difficulties distinguishing desires from other possible states with imperative contents such as pains. Those states didn't seem to involve endorsement/disapproval as the proper response. Using requests as the model for understanding the import of desires at least specifies the proper response that would then distinguish desires (and requests) from other kinds of attitude, like pain, intention, or decision. The engaged responses to desires mirror the responses to requests. In addition, desires seem to be conceptually connected with requests because the latter can be taken as natural expressions of desires in a communicative situation. Finally, the ability to utter and comprehend assertions and requests belong the linguistic competence of the folk which means that application conditions which make reference to speech acts should be epistemically accessible to them.

It seems, then, that one can hope to have a satisfactory explication of the concepts of belief and desire if their application conditions can be cashed out in terms of assertions and requests. The indispensability of the engaged responses to attitudes would then be inherited from the engaged responses to the corresponding speech acts that belong to the application conditions of the attitude-concepts. That idea is expressed in Christopher Gauker's communicative conception of beliefs and desires which I am going to analyse and defend in the next section. But there are some questions that need to be answered before doing that.

One worry might be that responses to speech acts, taken by themselves, don't completely match up with responses to the respective attitudes. Bernard Williams, for instance, has argued that there is a crucial difference between responses to assertions and beliefs. In the case of the latter their falsity motivates a "fatal objection", namely, that the person should give up her belief. Yet there are numerous circumstances where it is quite acceptable to utter false assertions. It is rather the insincerity of an assertion that matters for evaluating it as incorrect and this requires having certain beliefs and intentions backing up the assertion (Williams 2002: 70). If a person sincerely asserts that *p*, she also believes that *p*. In such a case she *should* retract her assertion when it is false because the aim of the assertion in this case is to share a belief with the hearer. I don't think that this is particularly damaging to the attempt to explicate the concept of belief in terms of assertions. The difference in responses doesn't cancel out important similarities. Agreement and disagreement are proper reactions both to assertions and to beliefs.

Nevertheless, following the previous problem one might raise a further and a seemingly more serious objection to *any* attempt to understand the application conditions of attitude-concepts in terms of speech acts. Namely, one might raise the concern that this similarity between attitudes and speech acts doesn't reveal anything interesting about the folk psychological practice. According to such an objection, the significance of speech acts such as assertions and requests should be explained in terms of the beliefs and desires that underlie them, not the other

way around. After all, assertions are supposed to express beliefs and requests to express desires. By trying to say something substantial about mental state attributions in terms of speech acts, the objection could go, I seem to put the cart before the horse.

The first thing to say in response is that one should acknowledge that there need not be a one-way understanding of assertions in terms of beliefs but beliefs can also be understood in terms of assertions (Williams 2002: 81). The same considerations should also apply to desires and requests. In the present context I take the analysis of the (attribution of) attitudes in terms of speech acts to be useful because it brings out the way in which we can explicate the concepts in an affectability-sensitive manner. Nevertheless, it could be objected that I have to make the controversial assumption that speech act production and comprehension is possible without the attitude attribution. If this assumption doesn't hold then the application conditions for mental state concepts can't be cashed out in terms of speech acts. The explanation of the latter would refer to the attitudes and this makes the analysis circular.

But is that circle vicious? If the aim is only to illuminate a certain aspect of the belief and desire attributions by appealing to speech acts, this doesn't entail that there can't also be a reverse way to understand speech acts in terms of attitudes. Williams' point, after all, is exactly that the two approaches can be complementary, not that there is a one-way understanding of the relation between attitudes and speech acts. Since the engaged responses to the latter are somewhat more apparent in our linguistic communication the present choice is to understand the less apparent responses to attitudes as similar to the ones we have to the speech acts. Nevertheless, it is possible to maintain this idea and to argue that we can't really understand or even produce speech acts without attributing mental states as well. The understanding of them is intertwined, so to speak, and the present decoupling is only for analytic purposes.

I do think that denying the viciousness of the circle is a feasible path to take here. If we concentrate on our actual situation, the understanding of attitudes and the understanding of speech acts are strongly linked. But for those who are still not convinced, one can also argue that speech act comprehension without mental state attribution is at least empirically possible. Imagine, for instance, a "ritualistic" community where discursive acts only follow pre-given rules that define the success of communication, independently of subjective interests and perspectives of the participants. In that kind of community, language can still serve the function of providing information and coordinating action. One could object that agents there would be irrational but this is beside the point. A more relevant objection could be that the content of linguistic utterances is ambiguous and identifying the beliefs and desires behind such utterances is required to disambiguate the message. One needs to make a stronger claim, though, in order to make this into a real objection. One should claim that ascribing beliefs is *necessary* for disambiguating messages. But as long as there are other means of disambiguation available or as long as there are messages that are not ambi-



guous, this objection has no bite. The burden of proof against these possibilities is on my imaginary opponent.

What's more, there *are* models of linguistic communication that don't necessarily require agents to have mentalizing abilities. For instance, Richard Breheny (2006) has defended a minimalist account according to which basic communication only presupposes the capacity to jointly attend shared situations. From another angle Ruth Millikan (2005) has argued for the feasibility of communication that is enabled by linguistic items which have conventional semantic mappings. Comprehension of these mappings doesn't necessarily require guessing at the propositional attitudes behind their production. The third example is Tad Zawidzki (2013) who takes linguistic communication to consist of understanding what is rational and proper to utter in a context considering the observable behaviour of others, without having to attribute propositional attitudes. According to Zawidzki, different mindshaping practices have caused us to have uniform expectations towards specific types of speech act as patterns of behaviour. Our responses to those speech acts can be explained without invoking our knowledge of the mental states that are supposed to underlie them.<sup>63</sup> I won't delve any deeper into the question as to which of these three accounts is the most promising because this would be a topic for an entire additional research project. In the present context I only mention these theories in order to strengthen the case for the intended explication.

There is also empirical support for the possibility of speech acts without attitude ascription. This comes from the fact that there are high-functioning autistic individuals who have a theory of mind deficit but are able to participate in linguistic communication (Andrews 2002). Even those who are able to pass the (explicit) false belief task show impairments in understanding mental states in real life social situations (Andrews and Radenovic 2006). The case of autism, then, indicates that there are people who comprehend speech acts but who are not competent attributers of beliefs and desires. This isn't a completely uncontroversial point because autistic individuals have also problems with mastering certain aspects of language, especially its pragmatic features. For instance, they have difficulties understanding reversibility of pronouns (Tager-Flusberg 1994), context-sensitivity (Happé 1997), metaphor (Rundblad and Annaz 2010) and irony (MacKay and Shaw 2004). The fact that autistic speakers are not fully able to master the intricacies of linguistic communication, however, doesn't mean that they don't grasp the difference between indicative and imperative utterance. For this, it suffices to be able to differentially respond to those two kinds of utterance, with agreement/disagreement and endorsement/disapproval, respectively.

---

<sup>63</sup> Zawidzki concentrates mostly on the question about the phylogeny of language, and speculates how the complex syntax might have its origins in the practices that involve costly signals, such as ritualistic displays (Zawidzki 2013: 167)

What about implicit false belief tasks? If people with autism were able to pass them, this would indicate that they still have theory of mind abilities and that their difficulties with the explicit tasks require a separate explanation. As it turns out, however, they don't manage to pass the implicit task either (Senju *et al.* 2010). I take it, then, that there are people who are able to participate in linguistic communication but are unable to attribute beliefs and desires. This indicates that not only is linguistic communication without propositional attitude attribution possible, there are also actual cases of this.

One could still object that it doesn't really amount to a *proper* understanding of a speech act if one doesn't take it to express the respective attitude. This objection seems unjustified because it simply excludes the possibility which I was arguing for. However, I can even allow that assertions and requests need to be explained in terms of attitudes and speak about quasi-assertions and quasi-requests instead. As long as there can be speech acts which have agreement/disagreement and endorsement/disapproval as their response profiles and which can be understood without attributing attitudes, they can be employed to explain a function of belief and desire ascriptions that I am advocating. Whether one calls them assertions and requests is merely a terminological issue.

The worry that the comprehension of speech acts requires mental state attribution, then, can be assuaged in three ways: first, by casting doubt on whether there's a vicious circle when we appeal to speech acts in our explanation; second, by considering a hypothetical linguistic community wherein speech acts are performed without mental state attribution; third, by pointing out feasible empirical models of linguistic communication that don't involve mental state attribution as a necessary element. Construing mental state attribution in terms of responses to speech acts is thus possible regardless of whether in our linguistic practice they are strongly intertwined. We can find such a construal in the form of Christopher Gauker's communicative conception of beliefs and desires to which I will turn next.

#### **4.4 The communicative conception**

As already mentioned, the communicative conception of propositional attitudes can be found in the work of Christopher Gauker. He defines it as follows: the paradigmatic case of ascribing a desire to someone is equated with commanding (or requesting) on her behalf and ascribing a belief with asserting on her behalf (Gauker 2003a: 221). For instance, when I claim about Mary that she believes that it is raining, I assert on her behalf: "It is raining". If I claim about Mary that she desires that her friend passes his exams I request on her behalf: "Pass the exams". Gauker himself illustrates the attribution of desires with the following hypothetical story. A house is being built. One of the builders, Balam, wants some rope and orders his assistant, Namu, to get some. Namu can then tell the keeper of the supplies that Balam wants some rope. In doing that, he actually passes over Balam's request (Gauker 2003a: 222). The example of a belief-

ascription also involves members of the same fictive tribe: Namu needs to know where logs are located and the leader of the loggers, Hanul, tells him that they are at the south fork of the river. Namu can then go back to other builders and attribute to Hanul a belief with the same content. This can also be described as an assertion on Hanul's behalf: "The logs are at the south fork of the river". The attribution of mental states in both examples can be analysed as the attribution of corresponding speech acts on behalf of others.

It can be argued that the previous examples don't show what is special about beliefs and desires. Namu could also have made his reports concerning Balam's behaviour in terms of assertions and requests not beliefs and desires. The choice between these possibilities seems to be arbitrary. Gauker makes a crucial qualification here: attributions of beliefs and desires are special because they do not have to be directly related to the actual assertions and requests of another person (Gauker 1994: 273; 2003a: 224)<sup>64</sup>. Gauker can maintain that although their attribution has its beginnings in the communicative situations depicted in the given examples, this practice has gained a relative independence from the speech act attribution. In order to attribute to someone a belief, it is not necessary that she has actually made a corresponding assertion. The same applies to the attributed desire. Beliefs and desires are in that sense more independent of the linguistic behaviour of the attributees than assertions and requests. For example, when I don't want to commit myself to another's claim or when I don't regard her as trustworthy, it is appropriate for pragmatic reasons to say that the other believes that claim but not to assert it. All in all, for Gauker, the differences between the attribution of speech acts and corresponding mental states reduce to what is appropriate to ascribe in a context and what's not. Sometimes the attribution of a speech act is appropriate while the ascription of a corresponding mental state isn't, and *vice versa*.

The communicative conception exemplifies nicely the proposal suggested in the previous section. There I proposed to explicate the application conditions of the concepts of belief and desire with reference to speech acts so as to make the appropriateness of engaged responses to the attributed attitudes intelligible. In Gauker's account, the ascription of a belief involves a vicarious assertion. Because of that the question of whether to agree or disagree with the belief comes naturally to the ascriber. Agreement/disagreement is a natural response to an assertion and also to a belief because of that. The ascription of a desire, on the other hand, involves treating the other person as if she had made a request. Requests are speech acts that are endorsed or disapproved of. Since the application conditions of the concept of desire make reference to requests, the question of whether to endorse or disapprove of the attributed desire is an intelligible move in response. The communicative conception treats beliefs and desires as having an inherent communicative import to which the possibility of

---

<sup>64</sup> This is a bit in tension with Gauker's claims about privileged access to one's own mental states, namely that person's willingness to assert that P is a good guide to his believing that P (Gauker 2003a: 234).

engaged responses is indispensable. The interpreted persons are treated by the interpreter as (silent) participants in the conversation (Gauker 2003a: 226) and the attributed attitudes have the same communicative significance as speech acts. The communicative conception is thus well-suited to explain the pertinence of engaged responses to attributed attitudes by referring to (vicarious) speech acts in the application conditions of the respective concepts.

It is probably quite counter-intuitive to equate the attributions of desire to X with requests on X's behalf. Requests are made to an audience but one can have desires that are not addressed to others. If a person wants there to be peace in the world does it make sense to say that it is appropriate to request on her behalf that there be peace? Who is this request directed at? Yet one can amend Gauker's conception by replacing requests with a more general category of evaluative utterances. Such utterances have the form "It should be the case that *p*" (cf. Van Cleave & Gauker 2010, 320). If we adopt this idea, the attribution of a desire that *p* should be taken as a claim on other's behalf that it should be the case that *p*. Here, I am still using the term "request" to mark such an utterance but it is important to bear in mind that the term can take both the second- and third-person form. However, if someone finds this use of the word "request" hard to stomach, she may simply think about evaluative utterances. Construing desire-attribution on the model of evaluative utterances still explains why the responses of endorsement/disapproval are appropriate.

Where in the philosophical landscape is the communicative conception located? First, the communicative conception is definitely not a version of behaviourism because mental states are not equated with dispositions to assert (Gauker 1994: 275). The latter aren't included in the application conditions either. For instance, it is sometimes appropriate to ascribe a belief to someone despite the fact that she lacks the disposition to assert the content of that belief. Gauker himself opposes what he calls "postulationism" according to which beliefs and desires are theoretical entities which are postulated in order to explain and predict behaviour (Gauker 2003a: 216). It isn't entirely clear exactly what kind of philosophical position can be identified with the postulationist conception. Functionalism is probably Gauker's primary target because he characterizes postulationism as the view that beliefs and desires are inner mechanisms that mediate between sensory inputs and behavioural outputs (ibid. 215). It is functionalism that analyses mental states in terms of their causal roles<sup>65</sup>. But the communicative conception also seems to clash with the identity theory which equates having mental states, either types or tokens, with having certain physical structures in one's head. The identity theory requires from the successful attribution of mental states more than simply that they be appropriately ascribable to a person. Because of its metaphysical commitments, identity theory is more demanding when compared with the communicative conception. All in all, the communicative conception is opposed to views that

---

<sup>65</sup> I am not going to delve into the differences between various functionalist positions because it's inessential for the present purposes.

try to find a deeper fact of the matter which beliefs and desires are supposed to be dependent upon or identified with. In that sense, the communicative conception has strong affinities to those strands in the philosophy of mind that are labelled “interpretationism”<sup>66</sup> In his earlier book Gauker also adopts this designation and calls it: “an account of what beliefs are need not be anything over and above an account of their attribution” (1994: 293).<sup>67</sup>

What distinguishes the communicative conception from other interpretationist positions is the way it depicts the function of folk psychological interpretation. As I’ve already mentioned, Gauker denies that the attribution of beliefs and desires with the aim of explaining and predicting others’ behaviour is the most fundamental form their ascription takes and instead stresses their role in linguistic communication. This brings him into conflict with Dennett, for instance, for whom the most notable value of having our folk psychological vocabulary is tied to the prediction of behaviour (Dennett 1987: 17). As far as I’m aware, the communicative conception is the only account of attitude attributions that let’s us see how the central function of these attributions is to enable the engaged responses that characterize affectability.<sup>68</sup>

In a later chapter I will come back to the issue of interpretationism and what the ontological commitments of the communicative conception are.<sup>69</sup> Here the focus is on the question of whether it can explicate the application conditions of the attitude-concepts in a satisfactory manner. As already noted, it seems to be in a good position to provide us with application conditions that explain why affectability is so central in the case of folk psychological attributions. In the next section I am going to inquire into whether it can also explain the other benefits of those attributions.

## **4.5 The communicative conception and the other benefits of folk psychology**

The denial that beliefs and desires are causally efficacious inner states and the seeming disregard of the predictive and explanatory roles of mental state attribution might leave a wrong impression. It might seem that the communicative conception doesn’t explain why we sometimes do manage to explain or predict the behaviour of persons in terms of propositional attitudes. For instance, let’s take a case when someone chooses to apply for a certain job in one company instead of another. It is natural to explain her choice in terms of her desire to have a large salary and his belief that the first company offers it. In

---

<sup>66</sup> For a clear articulation of interpretationism, see Child 1996: 23.

<sup>67</sup> In a later chapter I am going to argue against the interpretationist understanding of the communicative conception, but at that point I have made some crucial modifications to Gauker’s original account.

<sup>68</sup> Gauker doesn’t express himself in such terms.

<sup>69</sup> It should be kept in mind that this will be the communicative conception in a substantially modified form.

addition, if we don't know yet which job she is going to apply for, knowing her desire for a large salary, and her belief that the first company offers it, allows us to predict that she is going to apply for the first offer.<sup>70</sup> These examples seem appropriate. An opponent of the communicative conception could even use them as fodder for his argument that the attribution of desires and beliefs should track causally efficacious inner states. How else could we explain the success of those explanations and predictions if mental states weren't causally efficacious inner states? Already in the first chapter I stressed that the predictive and explanatory roles of mental state attributions can't be denied. Any account of the application conditions of the concepts of belief and desire needs to reckon with that fact.

Still, Gauker notes correctly that the claim that the concepts of mental states are applied in explanations and predictions doesn't automatically carry with it any postulationist commitments (Gauker 2003a: 238). Not every successful explanation or prediction of behaviour demands that the terms that figure in them refer to any definite inner entities. For instance, the explanation and prediction of behaviour in terms of character traits doesn't need to be interpreted in those terms unless one is prone to hypostatizing traits. The concept of character trait identifies certain behaviour patterns but does not identify a definite inner entity which is supposed to be behind those patterns.<sup>71</sup> This still leaves open the more important question: when we *do* sometimes ascribe mental states in order to predict and explain behaviour, how is it possible if the attributions of these states are equated with mere vicarious speech acts?

Gauker's own solution to the problem of how we are to account for the explanatory and predictive role of belief and desire ascriptions<sup>72</sup> is to demonstrate how mental state attributions, defined as assertions and requests on behalf of others, can still be used for the purposes of explanation and prediction<sup>73</sup>. He basically construes explanation and prediction as attributions of inner conversations (assertions and requests)<sup>74</sup>. Let's say that we give a rationalizing explanation of why someone acted in a certain way by relating it to her beliefs and desires. In this case we interpret her action by requesting on her behalf that she would achieve a certain goal and an assertion or a set of assertions which say how to achieve it. Take the example of a person applying

---

<sup>70</sup> We *can* predict. This doesn't mean that we have to do that.

<sup>71</sup> The analogy here may be a bit confusing because the communicative conception doesn't claim that *beliefs and desires* are constituted by behaviour patterns. The analogy between mental states and character traits is certainly imperfect.

<sup>72</sup> Note that I don't separate here the explanatory and predictive functions. I did that in the first chapter. For the present purposes a sharp distinction between them is irrelevant.

<sup>73</sup> Actually his first step is to argue against psychophysical laws that are supposed to underlie the causal connections between mental states and which the interpreters are supposed to rely on (Gauker 2003a: 239).

<sup>74</sup> "...an explanation of behaviour in terms of beliefs and desires is a recounting or reconstruction of a conversation." (Gauker 1994: 277)

for a job, for instance. When we explain her choice by her belief that the company she chose offers the best salary then this explanation can be reconstructed so that we assert on her behalf: “This company offers the best salary”. When we attribute to her the desire for a large salary we can also describe it as a command on her behalf: “Have a large salary!” Although these re-descriptions may feel a bit awkward, they are not absurd. The fact that we can make assertions and requests on others’ behalf often provides us with an opportunity to anticipate and understand their behaviour, at least on the assumption that people usually act in accordance with the assertions and requests we take them to have performed (Gauker 1994: 278).

An objection could be raised that the re-description of folk psychological explanations and predictions in terms of requests and assertions distorts this context of mental state attribution. If successful assertions and requests seem to require a potential response from an audience and the explanations and predictions of individual behaviour don’t seem to include this requirement, why should we construe them in the way the communicative conception demands it? We *can* do this but one could nevertheless argue that explanations and behaviours which lack a communicative context couldn’t passably be described in terms of assertions and requests. Let’s take the example of a person applying for a job. I’ve already admitted that the re-description of the explanation of his behaviour in terms of a vicarious request sounds awkward. But is it even appropriate given that an audience for the request is lacking?<sup>75</sup> The same could be asked about the belief that one company offers a larger salary and the corresponding assertion with that content.

I have two responses to this objection. First, there are self-directed assertions and requests. I can surely assert to myself “Uku, you’re behaving inappropriately” or “I have a meeting tomorrow” and give myself a request to behave better or to attend the meeting. These self-directed assertions and requests are also acknowledged by Gauker (1994: 274) and their feasibility is crucial for the communicative conception to work. In the cases when the audience seems to be lacking, a person can treat herself as the audience of her assertions and requests. When explaining the applicant’s choice between the two job positions by asserting and requesting on her behalf we can say that he also made the necessary assertions and requests to herself. In that sense, the re-description of mental state attribution in terms of assertions and requests isn’t necessarily a distortion of our explanatory and predictive practice.

Second, the re-descriptions of the explanatory and predictive mental state attributions in terms of assertions and requests on another’s behalf still gets the point of those attributions right. Asserting that a certain job is more profitable than the other and requesting that one get the job with the best salary can both explain and predict one’s behaviour, at least unless the agent is irrational and without suffering the weakness of will. These kinds of cancelling condition,

---

<sup>75</sup> Although, see Gauker 1994: 276 for a claim that there can be commands that that “do not seriously call on anyone to do anything.”

known as *ceteris paribus* clauses, apply to any theory of common-sense explanation and prediction, which means that they don't pose a threat to the communicative conception. In order for assertions and requests on behalf of others to play the explanatory and predictive role, we need only assume that people quite regularly act on the basis of assertions and requests. We simply need to assume that we can use this generalization to anticipate and make sense of others' behaviour. Explanations and predictions in terms of assertions and requests map quite neatly onto explanations and predictions in terms of beliefs and desires without changing the goal of the attribution. It is still explanation and prediction we are aiming at in these situations.

The problem that seems to deserve further elaboration is how to explain the putative fact that people tend to act according to assertions and requests. It could be objected that if this explanation makes reference to beliefs and desires, the communicative conception will be in trouble. But note that I have already rejected the claim that the explication of the application conditions of the concept of belief and desire has to be reductive. It can allow non-vicious circles. What is needed to act in accordance with assertions and requests is an open empirical question. Although I pointed out some models in 3.2 that don't presuppose the ability to attribute attitudes in order to understand speech acts, I don't think that it takes away from the point of the explication even if this ability *is* required.

But can the communicative conception also account for the mindshaping effects of mental state attributions? Remember that the mindshaping thesis was about how the attribution of mental states serves to mould the behaviour and minds of people by forming expectancies which the actions and thoughts of the interpreted persons should line up to. I took mindshaping to be one benefit among others that folk psychology offers. The communicative conception doesn't explain this benefit straightforwardly because nothing in it implies that the attribution changes or possibly changes the person being interpreted. But it is consistent with the possibility that as long as the attributee is aware of those attributions she will adjust her behaviour and thinking in response to the attribution. I took mindshaping to be a contingent benefit of folk psychology and I think that it suffices to point out that the communicative conception doesn't exclude the possibility of mindshaping. In any case, the communicative conception is actually in a stronger position than postulationism to explain mindshaping. The reason is that postulationism (which treats mental states as causal postulates) has difficulties explaining how brute causal forces could make normative bear upon the subject. Assertions and requests, on the other hand, can be taken as presenting such demands. My awareness that others treat me as if I had requested or asserted something may motivate me to act in accordance with those virtual speech acts. This may happen even if I haven't made the corresponding speech act myself. As analysed in the previous chapter, the mediating factor here might be the social superiority of the attributer, for instance, but the list of relevant factors is left open. We can say, then, that the



communicative conception is in a position to explain how the mindshaping effects are possible.

Actually all is not well and good with the communicative conception. In the next chapter I am going to argue that Gauker's appreciation of the *explanatory* role of mental state attributions doesn't go deep enough. Here I simply point out that the communicative conception has resources to account for other benefits than the engaged responses. In the remainder of this chapter I instead investigate an additional positive consideration in support of the communicative conception.

#### **4.6 On behalf of the communicative conception**

The aim of this subchapter is to give some further support to the thesis that the attribution of beliefs and desires must make reference to corresponding assertions and requests. To do that I am going to reverse the argument that was brought up against the communicative conception, namely, that the latter isn't able to make sense of folk psychological explanations and predictions. I think that Gauker has been quite successful in responding to that objection, as shown in the previous section. Nonetheless, he hasn't used it to his advantage to give additional support to the communicative conception.

To achieve that, I intend to turn the tables by challenging the postulationist to account for the communicative role of mental state ascription.<sup>76</sup> The communicative conception is relatively successful in deriving the explanatory and predictive uses from the communicative ones by demonstrating that the first can be (hopefully) always re-described in terms of the latter. But the possibility of a reverse move hasn't been demonstrated by the opponent of the communicative conception. If we assert or request on someone's behalf, can our actions be re-described in terms of explaining and/or predicting her behaviour without any reference to the vicarious assertions and requests? What follows is not a conclusive argument to prove that it is impossible but I argue that there is at least a reason to doubt that one could do it. This is also meant to strengthen the point already brought out in section 3.2, namely, that the accounts that explicate the conditions of mental state attribution without reference to speech acts don't explain the communicative import (i.e. the possibility of engaged responses) of the attributed beliefs and desires. There are occasions of belief and desire attribution with such a communicative significance but in which the attributions can't be re-described in terms of explanations or predictions.

Let us consider some examples of the communicative attribution Gauker himself used. There was the example of Namu who attributed the desire to get some rope to Balam and by doing that requested rope on Balam's behalf. He also attributed to Hanul the belief that the logs were at the south fork of the river and by doing that uttered the corresponding assertion on Hanul's behalf

---

<sup>76</sup> Not only postulationist because it concerns anyone who doesn't mention speech acts in the application conditions of the concepts of belief and desire.

(Gauker 2003a: 222). How could an opponent of the communicative conception analyse these attributions in terms of explanation and/or prediction? It is out of the question that by speaking about Balam's desire or Hanul's belief Namu in any way predicts their behaviour because the two have already acted. Their possible future behaviour isn't relevant in those contexts, at least not directly. This means that the only way of arguing for the re-describability must come from possible explanations implied by those attributions.

At first glance Namu's vicarious speech acts don't seem to explain Balam's or Hanul's actions because Namu's aim is to transmit requests and assertions. Which kind of actions are we even talking about? Although the answer seems to be clear (i.e. that we are interested in Balam's request and Hanul's assertion) I brought it up because it is not entirely evident why we should ask this question in the first place. After all, the explicit communicative context of those examples doesn't concern explaining the actions. It concerns requests and assertions getting conveyed and the identification of those speech acts makes a reference also to the possible responses of the audience (Hornsby 1994).<sup>77</sup> Merely asking what actions are being explained in those examples distorts the context of the attributions. The context is communicative, not explanatory. It is distorting to see the ascriptions of belief and desire as implicit explanations of the corresponding speech acts instead of as vicarious assertions because this would miss the point of the attribution. The point is to convey those assertions and requests and to react to them properly. Not treating the attribution as a speech act that is being conveyed would change the goal of the attribution.

I am actually going to argue in the next chapter that belief and desire attributions involve implicit explanatory commitments. But even if we grant that, the present challenge on the opponent is to show that the communicative significance could be fully described in terms of explanations. Only then would it be possible to claim that the component of asserting or requesting on others' behalf in mental state attributions can be reduced to some sort of explanatory activity. It isn't clear how this challenge could be met. Vicarious speech acts express certain commitments to the truth (assertions) or worth (requests) of the state of affairs that are being represented. Whether they involve explaining anything is left open by these commitments.

This irreducibility of communicative significance could also be expressed in terms of proper responses to it. Folk psychological explanations can call for the ascriber to make a judgement concerning the rationality of behaviour and thought that is being explained. The proper response to the communicative significance of belief and desire attributions is not that kind of judgement but agreement/disagreement or endorsement/disapproval which are directed at the attributed attitude. The response isn't to the attributed attitude's connection with behaviour and other attitudes. That we make assertions and requests on behalf

---

<sup>77</sup> One could speak here about explanations in a minimal sense so that any (true) report of someone doing something counts as an explanation because it makes. But we are considering here postulationist explanations which are not treated in such a minimal way.

of others while attributing beliefs and desires makes such responses understandable. The attempt at re-description in terms of explanation simply misses the significance of the belief and desire attributions in such contexts.

Before concluding this chapter I need to consider one more objection. This might discredit the idea that the communicative belief and desire attributions, exemplified by Gauker's story of the builders, are *mental state* attributions in any substantial sense. The phrase "I believe that..." can be used as a mere hedge to qualify one's statements without actually self-ascribing a belief. It might be objected in a similar manner that vicarious utterances which don't have explanatory or predictive import do not amount to belief- and desire-ascriptions *strictu sensu*. They could be seen as mere vicarious utterances, functioning as mere hedges to the conveyed speech acts. Gauker's own examples might be a bit unfortunate here because in the case of Namu, Balam and Hanul the supposed attitude-ascriptions boil down to attempts to convey others' speech acts. In such cases the hedge-interpretation might seem pertinent. If we concentrate only on Gauker's examples, then, the present objection might have a punch. But it is important to note that the appropriateness of attributing a speech act and of attributing the corresponding attitude don't coincide in Gauker's conception. There are occasions when one can ascribe the attitude but it would be false to say that the attributee has made the corresponding utterance and *vice versa*. Nevertheless, what matters in such cases is still the communicative significance of those attitude ascriptions which can't be reduced to their explanatory role. For a postulationist this qualification might not suffice because it doesn't correspond to her intuitions about the causal commitments of the mental state attributions. But note that we are here explicating the concepts of belief and desire and the explication might deliver application conditions of those concepts that need not match the intuitions of the ascribers. The communicative conception is justified as long as it helps us understand how folk psychological attributions imply the indispensability of engaged responses and as long as it explains how it is possible to explain, predict and mindshape on the basis of belief and desire attributions.

Nevertheless, this objection has its point. In the next chapter I will argue that the original communicative conception overlooked the importance of the explanatory benefit of belief- and desire-attributions. This means that it should be complemented in a crucial respect. Here I only wanted to show that Gauker's account can be taken as providing the initial explication of the application conditions of the concepts of belief and desire. The reference to speech acts seems to be a necessary component if those conditions are to explain the indispensability of affectability. The mental state attributions can't be understood merely in explanatory and predictive terms.

## 4.7 Conclusion

In this chapter I argued that if we take affectability to be the central benefit of folk psychology then the communicative conception of beliefs and desires is the most natural way of construing the conditions for attributing those attitudes. Since the attitudes share their communicative significance with speech acts, modelling mental state attribution in terms of the latter helps us to see why the engaged responses to the attributed attitudes are appropriate. The next chapter investigates the communicative conception further and raises some doubts concerning its inability to fully account for the benefits of mental state attributions. Although I am going to argue that the original communicative conception can't fully account for the application conditions of belief and desire, I maintain that the assertions and requests on behalf of others constitute a necessary condition for the attribution of beliefs and desires.

## **5. MODIFICATION OF THE COMMUNICATIVE CONCEPTION**

### **5.1 Introduction**

In the previous chapter I defended the communicative conception of beliefs and desires as the most promising explication of the concepts of those attitudes. It enabled us to see the appropriateness and indispensability of engaged responses to attributed attitudes. Gauker's initial account of the communicative conception provided the necessary conditions for the attribution of beliefs and desires. This chapter will take a critical stance. I will argue that the original communicative conception has to be elaborated in order to do full justice to the benefits that mental state attribution delivers. The aim is to show that there is a genuine explanatory need that the attribution of mental states satisfies and which needs to be accounted for by the communicative conception. The further aim is to modify the communicative conception in a way that does justice to that need.

I will take it for granted that the attribution of beliefs and desires involves assertions and requests on the behalf of others as was established in the previous chapter. The problem with the communicative conception, I claim in section 5.2, is that, while it accounts for the affectability of mental state attributions, it fails to explain why the concepts of belief and desire are needed in addition to an understanding of assertions and requests on others' behalf. The accounts of both concepts need quite substantial alteration in order to be satisfactory. After having argued for that, in section 5.3 I will present the modified communicative conception. According to it the necessity of postulating desires arises in situations wherein the requests and hedonic states of persons don't match. As for beliefs, I claim that they need to be postulated in cases when the dispositions of persons are in tension with their actual assertions. In section 5.4 I draw some comparisons between the modified communicative conception and some other approaches to the concepts in question. Finally, in section 5.5 I will consider some possible objections to my account.

### **5.2 Problems with the communicative conception**

#### **5.2.1 The initial worry**

The communicative conception takes attitude attributions to be speech acts that can be performed in specific situations. But one may feel (very) uncomfortable about the claim that the difference between beliefs and assertions (also desires and requests) boils down to a pragmatic distinction. Isn't there some essential disparity between them? This intuitive shortcoming of the communicative conception isn't yet an argument against it. Gauker could claim that he aims at providing the most coherent theory of beliefs and desires which takes into

account only a subset of intuitions people have on the subject. In the previous chapter we saw that it does seem to explain the main benefits of folk psychological attributions. One would need to elaborate on the feeling that the attribution of beliefs and desires doesn't merely consist of the making of assertions and requests on the behalf of others.

I do think that there are various points where it is possible to raise some serious concerns that go a bit further than merely stating the intuitive point. It is safe to say that Gauker doesn't merely want to fit the attribution of beliefs and desires somehow or other into his overall theory. The communicative conception should still be about the concepts of belief and desire that people actually use. What could be the evidence that Gauker has or has not managed to achieve that? From the perspective of traditional conceptual analysis, the answer seems to be simple – if the proposed conception of beliefs and desires doesn't capture the most important intuitions one has about those concepts then it isn't satisfactory. Since in the present framework the aim is not to provide an analysis of those concepts that is supposed to respect *all* intuitions I am blocked from taking that track. I still think, though, that we can sometimes say when a theory doesn't actually explain what it aims to, for the reason that it explains something else. This is a very difficult issue, of course. The most noteworthy example of such difficulties is perhaps the discussion about how to explain consciousness.<sup>78</sup> Nonetheless, there do seem to be unproblematic contexts wherein the evaluation of a certain explanation as unsatisfactory (in the sense that it misses its *explanandum*) is reliable. This doesn't require a full analysis of the *explanandum*.

So what is it about the communicative conception that gives us reason to doubt that it actually captures our everyday notions of belief and desire? The first suggestion is that the idea of attitudes as devices for performing vicarious speech acts doesn't actually reckon with the fact that when we say that someone believes or wants something we say something about that person. For instance, we seem to say something about her subjective perspective on things. This suggestion doesn't yet amount to a criticism of the communicative conception. Nevertheless, for the moment it suffices to say that the conception of attitudes which don't express their possessor's subjective perspectives makes such attitudes unrecognizable to us. It does so in the same way that the attribution of attitudes without engaged responses made them unrecognizable.

My critique of the communicative conception in this chapter concentrates on a more concrete difficulty. Remember that according to Gauker's communicative conception the concepts of belief and desire were adopted as a means of transmitting the speech acts made by other people. He also illustrated this with a hypothetical story about builders conveying assertions and requests to one another which were then labelled "beliefs" and "desires" (Gauker 2003a: 222).

---

<sup>78</sup> Take Dennett for instance, in case of whom the critics of his theory of consciousness (Dennett 1991a) have often argued that it actually isn't about consciousness (see Chalmers (1995), for the objection along these lines).

The problem we found with that suggestion was that it didn't seem to reveal any discursive need that was over and above the need to speak about assertions and requests. It really doesn't add anything besides new terms to a linguistic practice. These terms simply coin the assertions and requests that are already in use under the labels "assertion" and "request" with new terms "belief" and "desire".

Gauker's theory does have resources that enable us to adjust the aforementioned story in a way that would present a more distinctive role for the concepts of belief and desire. This is because assertions and requests on others' behalf form two specific kinds of discursive move that are responsive to a need which can't be satisfied by mere assertions and requests. There are occasions in which mental state ascriptions have an application that can't be fulfilled simply by ascribing a speech act. This can happen, for instance, when the attributee doesn't perform the corresponding speech act herself or when the attributer isn't ready to take the attributee as authoritative (Gauker 1994: 273). Let's call these two cases the occasion of *absent utterance* and *declined authority*. Keeping that in mind, the adjustments to the original account would be relatively simple. We simply have to consider situations wherein the agents find it appropriate to assert and request on the behalf of others without acknowledging their authority or wherein they haven't produced the corresponding utterance themselves. Having the attitude-concepts, then, fulfils the need to perform vicarious speech in such situations. Isn't such contextual usefulness all we need? The question is whether we can remain satisfied with that suggestion: do the situations of absent utterance and declined authority constitute a sufficient condition for certain things to qualify as beliefs and desires?

### 5.2.2 Conceptual need for mentalistic concepts

We have defined *absent utterance* and *declined authority* as those situations in which talking straightforwardly of assertions and requests on others' behalf are somewhat inappropriate. Introduction of the concepts of belief and desire seemed to be useful because it allowed us to draw the distinction between full-blown and qualified vicarious utterances. We can now ask whether this move by Gauker suffices to make the motivation for introducing these concepts explicit. I will argue that the communicative conception in its present form still doesn't manage to articulate a substantial conceptual need behind the attitude-concepts. To see that we should first consider what the notion of "conceptual need" actually amounts to.

A conceptual need isn't, strictly speaking, a psychological state. It is rather a state of affairs which is defined in terms of the absence of the concept in question. It arises when people find their conceptual repertoire insufficient for conveying something distinctive about the world or themselves. It can be appealed to when we are interested in understanding the main benefits that the application of the concept offers. If a proposed need can actually be fulfilled by existing means then we have reason to doubt that such a proposal is acceptable.

One can find a similar appeal to conceptual needs in the genealogical approach to concepts pursued by such authors as Bernard Williams (2002) and Edward Craig (1990). These accounts present us with a hypothetical situation in which the concept in question isn't yet in use. One can then ask what needs could be fulfilled by introducing that corresponding expression into the vocabulary of the community. This kind of speculation is intended to bring out the distinctive role of the concept that is under scrutiny and its connection to other concepts. The present question is whether Gauker has been able to do this with the concepts of belief and desire. In a way, then, the following articulation of the problem can also be taken to be a contribution to the genealogical accounts of concepts.

Before going on we should note that the concepts that are used in genealogical explanation are also linguistically articulated. The genealogy of a concept is supposed to articulate the need for introducing a new *word* into the vocabulary. Because of this, there won't be any strict distinction between the concept and the word that is supposed to express it. In what follows, I will talk about them interchangeably. That probably seems like a seriously troublesome move for many because words are specific to a language. The genealogical account of the concept of belief as the term "belief" faces the threat of explaining only the English word. But in this case it wouldn't say anything about the synonymous words in other languages that also denote beliefs (e.g. *Überzeugung* in German, *uskumus* in Estonian, etc.). It seems that if we want our account of the attitude concepts to be generalizable, we *should* make the distinction between words and concepts, and ask only about the latter.

I do agree that different words like *belief*, *Überzeugung* and *uskumus* have something in common and the idea that they express the same concept makes that commonality explicit. For the purposes of understanding the conceptual need for the concepts of belief and desire, I will assume that these concepts can be individuated in terms of the inferential significance of words. In that way the words in different languages can be taken to express the same concept.<sup>79</sup> When accepting this conception, the genealogical approach makes much more sense because we can ask about the function and significance of a concept by considering the role of the corresponding word in our language-use and conscious thought. What is discovered about the word *belief* will hopefully also hold for the words *Überzeugung*, *uskumus*, etc., which may differ in their phonemic properties but can share the same inferential-functional role (Sellars 1974: 425).

But does admitting this commit me to a specific theory about the structure of concepts? When I individuate concepts by the inferential roles of words, don't I simply exclude the possibility of an atomistic theory? We should keep in mind, though, what it is that we are presently doing. "Concept" is a term that is used in many different ways, for many different purposes (Machery 2009). Unlike

---

<sup>79</sup> This conception is quite obviously influenced by Bob Brandom's inferentialist approach to conceptual content, going back to Wilfrid Sellars (Brandom 1994). The difficulties with Brandom's theory – most notably, the problem of holism – are numerous but for the present purposes I will assume that the general approach is on the right track.



theories about conceptual structure, I am here not really interested in the nature of concepts as such. As long as the genealogy enables us to reach some understanding of the possible role and value of a concept for our self-understanding, it can be counted as successful. Furthermore, if this requires thinking of concepts as embodied in our language-use and embedded in inferential structures, then I see no reason not to take the endeavour to be legitimate. By exploiting the possibility of bringing words and concepts together through the words' inferential role, I remain agnostic on the question of whether there are other ways of understanding concepts that handle them in a way that is independent of language.

So what are those conditions which necessitate the introduction of a new concept? The communicative conception is a view that by attributing attitudes we perform certain communicative acts which count as such due to convention. I understand convention to be a set of assumptions governing some social practice. Under what conditions do we need to introduce a new term to label an element in such a practice? This brings us to a rather difficult issue because it isn't even clear if practices can be individuated without mentioning the mental states of the community members. Setting that worry aside, it is reasonable to claim at least that a minor change in a convention doesn't mean that the practice it pertains to or its particular elements need be renamed. A social category can remain constant through minor changes as long as it fulfils its function. For instance, after the presence of a priest's presence has been deemed unnecessary, marriage can still be called "marriage", the bridegroom can still be called "bridegroom" and the bride be called "bride". The conceptual needs for introducing terms for social practices and statuses are quite various but the introduction of a new term should have distinctive discursive and practical consequences. In talking about distinctive consequences I am relying on intuitive considerations but the following discussion should demonstrate how they can be put to use in order to evaluate Gauker's account.

Let's think now about the agents who have come up with the practice of making assertions and requests on each other's behalf. Consider those occasions in which the ascription of beliefs and desires (as opposed to the uttering of vicarious speech acts) seems to become necessary. According to the communicative conception, we have here two situations wherein the straightforward speech act on another's behalf can be deemed inappropriate – when the other person doesn't perform the corresponding assertion or request herself (absent utterance) and when she is not taken as authoritative (declined authority). The question, in a nutshell, is – should the agents in such situations adopt new concepts in order to distinguish these two kinds of vicarious utterance from those that are made when the respective utterance is *not* absent and authority *is* transmitted?

First, we should note that there's a problem with the suggestion that declined authority *and* absent utterance necessitate introducing a new concept. Assuming that both situations are cases of genuine conceptual need, the question arises: given that these two cases are very different, why should they motivate the

introduction of the concept of belief in the case of assertions and the concept of desire in the case of requests? It seems that if they both express genuine conceptual needs then they should give rise to two concepts, both in the case of assertions and in the case of requests. One would identify non-authoritative speech acts and the other would identify speech acts on another's behalf without that other having made the relevant utterance. These two concepts seem to diverge in their significance. The cases of declined utterance and absent authority, then, seem like entirely different cases for introducing the concepts in question.

So let's consider these two cases separately, starting with declined authority. Does this situation articulate the need to adopt the concepts of belief and desire? I am not sure if the answer has to be affirmative. The case of declined authority doesn't seem to express a genuine conceptual need at all because the fact that the attributer isn't sure about the attributee's authority doesn't alter the fact that the latter has performed the speech act. The subject simply conveys it. It would be unnecessary to invent a new label for this act. It would be sufficient to say: "She asserted that  $p$  but don't take her seriously"; and "She commanded that  $p$  but don't take it seriously."

Note that when I deny that the case of declined authority reveals a real conceptual need I don't base my claim merely on the fact that the sentence in which the new term is applied can be paraphrased into another sentence without that term. It is rather about asking what point there is to introducing the new term. The answer seems to be negative in the present case. Of course, it is useful to distinguish between authoritative and non-authoritative speech acts. Nonetheless, it actually seems to be less confusing when we draw this distinction while maintaining the terms "assertion" and "request" in both cases. If we maintain these terms then we won't lose sight of the fact that the utterance was also uttered in the case of declined authority. For instance, if a child tells her parent to bring her ice cream, reporting it as a command with a comment that it was merely a child who requested it, should be unproblematic.

It is possible to think up a situation wherein introducing a new term in the case of absent authority would have a more far-reaching significance. I don't deny that. For instance, one can imagine a community wherein those who perform unauthorised speech acts are sanctioned or even punished. In such a community, the introduction of a new concept would have distinctive practical consequences. However, imagining such a possibility requires us to make very particular (and rather far-fetched) assumptions about the community in question. Such a conceptual need does not explain the concepts that are used in the actual world. I take it, then, that the case of declined authority doesn't motivate the introduction of the concepts of belief and desire. It should offer a more substantial and stable benefit.

As for the case of absent utterance, the question is why do we need to speak about beliefs and desires in a situation wherein we are inclined to assert or request something on the behalf of others but wherein they haven't made the respective utterance themselves? Why not simply say that they *would* assert or

request it if the circumstances were more accommodating? Why does one need to coin new terms for such discursive situations? Doing it doesn't seem to answer real conceptual needs but simply invents labels for something that can already be expressed in terms of (potential) assertions and requests. When a new concept is introduced to represent a form of social practice the latter has to be genuinely new, with distinctive practical consequences. By labelling the dispositions to command "the concept of desire" wouldn't satisfy any genuine conceptual need. The same seems to apply to assertions and beliefs. So the initial response to this proposal is the same as in the case of declined authority.

To counter this problem Gauker might stress that the ascription of a mental state to someone doesn't even require that she be disposed to produce the corresponding utterance (Gauker 2003a: 225). But if the attributee need not even be disposed to perform the speech act that is uttered on her behalf then an obvious question arises: on what grounds we can perform that vicarious speech act? This shouldn't be an entirely arbitrary matter but Gauker himself doesn't provide any suggestions about the conditions under which the attribution would remain appropriate without the disposition on the attributee's part. Yet without these grounds, the vicarious speech acts seem to be ill-motivated. What's more, this proposal leaves us in the dark about what need there is to perform such seemingly baseless speech acts on another's behalf without the other even being disposed to utter those speech acts herself. In order to make this need explicit one has to say something more about the conditions under which such vicarious speech acts can be performed. Otherwise these practices (and the introduction of concepts to mark them) wouldn't make much sense. There should be a reason to make such an assertion or command.

Gauker might respond by making some further distinctions between different kinds of vicarious speech acts in order to account for the intuition that the attribution of beliefs and desires should say something about the grounds for performing speech acts on the behalf of others in the case of absent utterance. For example, he could distinguish between person-relative and non-person-relative speech acts. The first imply that the person on whose behalf it is made agrees with or endorses it but the latter leave this open. The first case would perhaps provide us with the condition under which the speech act on another's behalf is appropriate, even if she hasn't performed it herself. What's more, wouldn't the need to utter vicarious person-relative speech acts also suffice to account for the need to introduce new concepts? After all, it would allow the agents to relate vicarious assertions and requests to the subjective perspective of the attributees (as long as the readiness of the attributees to agree with and endorse such speech acts expresses their subjective perspective). This seems like a practically significant and distinctive move in the discourse.

This suggestion is actually relatable to my own positive proposal but it involves an ambiguity as it stands. The person-relative speech act on the behalf of another either requires that another person explicitly agrees with or endorses it or that she would agree with or would endorse it in certain circumstances. In the first case the need for introducing a new concept falls away – one could

simply attribute to her the assertion that *p* or the request that it should be the case that *p* – after all, she explicitly agrees with or endorses them. But if she merely *would* endorse them then one can always ask for the grounds on which she would do that. The answer to that question should say something about the person but the communicative conception leaves us in the dark about what it is. In any case, what the endorsement of the vicarious speech act amounts to needs further elaboration and Gauker hasn't done that. My own positive proposals will exploit this lacuna in his account.

My claim, then, is that the communicative conception doesn't manage to explain why one would need to coin new terms for use within the practice of performing vicarious speech acts, thereby arriving at the ideas of belief and desire. It either gives these concepts a communicative role which is too shallow to motivate the introduction of new terms or it faces the problem of explaining the grounds for performing speech acts, unable as it is to make intelligible the need for the concepts of belief and desire. This still isn't a conclusive proof against the communicative conception. One could maintain, for instance, that concepts of attitudes really do play such a shallow role. But an alternative explanation would have an advantage over Gauker's position if it demonstrated the existence of a more substantial conceptual need.

### 5.2.3 The solution?

The inability of the communicative conception to explain the need for the concepts of belief and desire poses a challenge. How should we proceed in order to remedy the situation? I suggest that we can mostly retain Gauker's main idea that the attribution of attitudes involves making assertions or requests on behalf of others but we need to complement it by reconsidering how it relates to the conceptual needs that the agents might have in the case of absent utterance.

The explanation for beliefs and desires that I have in mind starts with the following hypothetical situation. The agents have linguistic competence and the ability to perform assertions and requests. In that respect, it follows Gauker's lead. It also acknowledges the fact that those agents need to assert and request on the behalf of others and that sometimes this may happen in cases of absent utterance<sup>80</sup>. But I claim that the explanation of why a concept of desire is needed in such a community should also make explicit why it is proper to perform a speech act on the behalf of another even if the other didn't make the corresponding utterance herself. You need to have a reason to assert or request something on another's behalf. This means that there should be an explanation of why this kind of action is appropriate. It is natural to say that someone's having a belief or desire *explains* why it is appropriate to make a corresponding assertion or request on her behalf. The grounds for performing vicarious speech

---

<sup>80</sup> Since I've dismissed the case of declined authority as involving a genuine conceptual need, the focus will be on the case of absent utterance.

acts, then, should be the key to revealing the conceptual need for the attitude-concepts. It is not a mere detail that needs to be faced after the communicative conception has already been established.

Is this concession a return to postulationism which Gauker opposed? It certainly differs from his account in admitting that the concepts of belief and desire have a distinctive explanatory role. But since I haven't yet answered what kind of explanatory posits they are, it is left open whether beliefs and desires are causal postulates or not. I can also continue to agree with Gauker that the attribution of beliefs and desires doesn't have much predictive potential. Most importantly, I stick to the idea that the attribution of beliefs and desires involves assertions and requests on the attributee's behalf. But the two problems I've presented mean that one cannot remain as deflationist as Gauker does. The account of belief and desire attribution needs to say something about the conditions under which a person is properly describable in terms of such concepts. I don't exclude the possibility that assertions and requests on another's behalf could form the core of the concepts of belief and desire – just as Craig conceded that the justified true belief may be a correct but minimal definition for the concept of knowledge (Craig 1990: 33) – but this is only a starting point for a more substantial account.

## **5.3 The modified communicative conception**

### **5.3.1 Desires**

To see how the communicative conception needs to be complemented, let's start from the concept of desire. I hope that it is a relatively intuitive point that one can distinguish between mere acknowledgements of requests and actual attributions of desires. The latter should also take the perspective of the attributee into account. From the previous analysis we saw also that the most promising account of the conceptual need – that of person-relative speech acts – took attitude attributions to say something about the attributee. My proposal with regard to desire attributions is that the ascription of a desire opens up the question of whether the interpreted person will also be satisfied when the represented state of affairs is realized. The ascription of a request doesn't have such a consequence. By saying that another person orders or would order something one doesn't necessarily say anything about her actual sympathies. Gauker, of course, is more subtle in that the practice of making requests on the behalf of others is meant to be separable from the actual utterances of individuals. But as we saw in connection with the problem of grounds, Gauker doesn't provide many positive suggestions concerning the ascribability conditions of such vicarious requests. The alternative account, then, should make explicit how desire attributions explain when it is appropriate to request on another's behalf and relate this somehow to the attributer's understanding of the attributee's subjective satisfaction.

To see how to do this, let us now reconsider the hypothetical situation wherein the concept of desire isn't yet introduced but wherein the agents engage in the practice of making requests to each other. This is something that Gauker needs to allow in order to explain the need for the concept. It is reasonable to assume that such agents also have some primitive abilities for social cognition such as emotion-recognition and for understanding how people are affected by events: whether they are pleased or displeased; satisfied or frustrated. This claim is also plausible for empirical reasons (see Nichols (2001, 436) for the view that the understanding of affect doesn't require possession of the ability to attribute propositional attitudes). One quite obvious move to make, if our aim is to clarify what subjective satisfaction amounts to, is to take into consideration our capacity to recognize when someone is pleased.

Doesn't the admission of affective understanding already bring with it the understanding of desires? After all, it is quite natural to say that if someone finds something pleasurable she wants it. One can doubt, though, that the mere fact of something being disposed to cause pleasure to a subject amounts to her having a desire for it. What is pleasurable for a subject need not be desirable for her. What's more, from the present perspective, introducing a new term "desire" for something that can already be satisfactorily understood in terms of potential pleasure is pragmatically rather pointless.

Another important thing to note is that the understanding of pleasure doesn't imply that the agents in question have to be postulationists about it. They don't need to identify pleasure with some causally efficacious inner states. In order to understand that another person is pleased with a state of affairs, nothing more is needed than certain observable criteria that help one recognize that the other is in that condition. These criteria can include linguistic utterances, bodily expressions and temporally extended patterns of behaviour; anything that is relevant for recognizing that the person is satisfied. The present appeal to the understanding of affect, then, isn't opposed to the spirit of the communicative conception.

Now the question is what the concept of desire could enable those people to do that wasn't possible before. They understand evaluative utterances and are able to realize when a person feels pain or pleasure, is content or frustrated. In that kind of state circumstances will emerge in which a person who utters a request doesn't seem to be personally satisfied when it is carried into effect. There are also situations wherein she would presumably take pleasure in some state of affairs but hasn't requested it. The need to speak about desires arises when there is a tension between the objective satisfaction condition of a command and the subjective affect that it arouses. These are situations wherein the person's requests don't seem to reflect her actual needs. The ascription of a desire is essentially perspective-sensitive because it not only states a request on another's behalf but also relates it to the question of what really pleases that individual. In that way desires are still akin to requests but are also tied to the perspective of the agent, the explanation of which is lacking in Gauker's communicative conception. For instance, imagine a situation wherein I know

that another person really enjoys eating broccoli, we're in a situation where she is hungry and there is some broccoli in the fridge. Although she hasn't explicitly uttered a request for broccoli, I find it natural to ascribe a desire for broccoli to her, thus treating her *as if* she had requested it. In that way I can endorse such a (virtual) request and find means to satisfy it. I have extended the engaged response of endorsement which is characteristic to requests to the case of absent utterance.

The conceptual need that the notion of desire satisfies is, then, the need to know which requests would lead a person to her subjective satisfaction and which would not. In addition, this proposal nicely explains the grounds for making vicarious requests in cases of absent utterance. It does so by relating them to the ascribers' ability to anticipate and recognize others' pleasure. Gauker lacked the means to make explicit what the grounds for vicarious requests in such cases could be. The present account, on the other hand, provides a reason to treat the other person *as if* she had uttered a request because satisfying the latter would be pleasing to her.

It is also noticeable that this line of thought forces us to recognize an irreducible explanatory role played by the concept of desire. The modified communicative conception brings together Gauker's idea of the attribution of desire as a request on another's behalf and the idea that the mental state concepts play an explanatory role. But it is important to note that, as with admitting the understanding of affect into our account, this concession doesn't imply postulationism about desire attributions. The latter would be the case if the explanatory role of desires were to track causally efficacious inner states. In the present case, the explanation merely makes intelligible why the vicarious utterance is appropriate. As already noted, the recognition of someone's being pleased doesn't require postulating inner causes behind a person's expressive behaviour. The desire attribution only requires for its success that the attributee would be satisfied if the claim on her behalf were fulfilled. That can be understood on the basis of overt criteria.

One might wonder whether my proposal encounters a difficulty that is similar to the one faced by Gauker's initial account; namely, whether the need to understand which requests would please the attributee is really such a substantial conceptual need for the concept of desire. It could be argued that introducing a new term into such a situation is unnecessary. Why not simply speak about dispositions to feel pleasure in certain situations? The key here is that vicarious claims still play a role in the present account. Since my own proposal retains Gauker's idea that the desire attributions involve requests on another's behalf the concept of desire can't be reduced to the concept of a disposition to feel pleasure. It articulates a connection between speech acts and affect that helps its users achieve something distinctive in their linguistic practice. Note that this claim about distinctiveness doesn't appeal to the impossibility of paraphrasing desire-ascriptions into vicarious commands conjoined with pleasure-attributions. The paraphrase might be possible just as it was possible to talk about qualified command-reports instead of desires in the

case of declined authority. The distinctiveness in question concerns what desire attributions enable us to do. They have a synthetic function. They articulate the connection between two kinds of phenomena: imperative speech acts and affective conditions. What's more, they explain why it is proper to extend the practice of commanding to the cases of absent utterance. The original communicative conception didn't extend the use of the concept beyond the domain of speech acts and didn't assign any explanatory role to the concept. The present proposal, on the other hand, does both.

Let's elaborate on the merits of the present account. Explaining the concept of desire in terms of pleasure also connects it naturally with its wider functional significance in human social life, thus demonstrating the explanatory potential of the modified communicative conception. Having the concept of desire as a tool for connecting the communicative behaviour of agents with their affective expressions enables people to think about whether making claims on another's behalf will actually be advantageous to her when the command gets fulfilled. This opens up specific ways of coordinating with others. The best way to show this is to consider how the cooperative actions would look if people didn't attribute desires to one another. Such actions would still have goals in relation to which the behaviour of individuals could be evaluated. The imagined community could still be engaged in building houses and shelters, gathering and growing foodstuff, educating children, performing rituals, etc. During these activities people would command each other to do things and also to command on behalf of one another. All of this fits into Gauker's account alongside the fact that the mere practice of vicarious requests doesn't require the introduction of the concept of desire into the communicative practice. What that imagined community lacks is the means of connecting the question of whether the practice of commanding is successful with the question of whether participants of that practice will be subjectively satisfied with the particular claims made.

This is basically a repetition of my earlier points. But these points are important to bear in mind if we are to recognize the full significance of having the concept of desire. Introducing that concept into the aforementioned situation enables people to see the success of their communicative practice in a new light. Now the success can be evaluated not only on the basis of the fulfilment of objective goals but also on the extent to which individual agents are satisfied with it. Taking the idea that people have desires seriously might even cancel the objective criteria for evaluating the success of the practice. At least the desire-based criteria might compete with the objective ones. Assuming that people are somewhat altruistically motivated, at least to their ingroup members, the need to know whether the communicative practices of a group bring satisfaction to its members is important. Employing the concept of desire enables us to satisfy that need. In that sense this concept is important for leading a worthy communal life and for coordinating action.

But does the concept of desire also open up new ways of manipulating others? I don't deny that at all. For example, one can deceive others into thinking that a person wants some state of affairs to be realised even though the



attribution isn't actually sensitive to the affective profile of that person but instead only to that of the subject/attributer. Let's consider a situation wherein another person is more authoritative in her wishes than the attributer. In that case the latter might achieve what she wants by putting others to work on the false assumption that the person whose vicarious commands they are satisfying actually wants them to be satisfied. However, in actual fact, they are serving the subject who has attributed such a desire to the more authoritative person. By exploiting the fact that desire attributions are supposed to bring together the requests on another's behalf and the other person's satisfaction, one can deceive people into mistakenly seeing such a connection. The present example is only one among many strategies to do that. It is almost needless to say that these manipulative moves undercut the original purpose of the desire attributions. If people always decoupled vicarious commands and subjective satisfaction from one another then the practice of attributing desires would presumably lose its point.

The modified communicative conception, then, handles the intuition that desires involve subjective satisfaction very well. It makes explicit the functional significance of the concept of desire and offers an explanation of the cases of absent utterance. In all these respects it is superior to Gauker's original account. Nevertheless, it maintains that the attribution of desire involves requesting on someone's behalf. It also avoids the postulationist claim that the attribution is involves reference to the inner causes of people's behaviour. The addition of pleasure-recognition makes the account perhaps less deflationist but that is the price we have to pay if we want to hold onto the main insights of the communicative conception.

I now move on to account for the concept of belief and for the conceptual need to adopt it.

### 5.3.2 Beliefs

As in the case of desires and commands we also distinguish between beliefs and assertions as distinct categories. A person asserting that  $p$  isn't the same as her believing that  $p$ . Nor does a person believing that  $p$  necessarily mean that she asserts or is disposed to assert that  $p$ . Of course, the communicative conception as presented by Gauker doesn't identify beliefs with assertions or dispositions to assert. Instead, according to Gauker, the need to assert on the behalf of others is the source of the concept of belief. Yet the criticism made above applies also in the case of beliefs. There seems to be a deeper conceptual need involved in talk about beliefs than merely the need to make assertions on each others' behalf. The latter wouldn't necessitate the introduction of a new term. As with desires we should consider the explanatory role that the concept of belief could play in the practice of asserting on the behalf of others.

My proposal for the concept of belief is somewhat different from the account of the concept of desire. In the case of the explanation of desires the subjective satisfaction conditions were explicated in terms of actual and potential hedonic states. The explanation of the concept of belief doesn't appeal to anything like

that. Beliefs don't have subjective satisfaction conditions like desires.<sup>81</sup> What counts is whether they are true or not.<sup>82</sup> Yet the case of absent utterance is still important if we are to understand the significance of belief ascriptions and the latter should still say something *about* the attributee.

If we consider the possible needs for making such seemingly groundless vicarious assertions it might be of help to think back to the engaged responses of agreement and disagreement that constituted the affectability of belief attributions. Situations arise wherein a person hasn't made an assertion that *p*, yet she behaves in way which is in accordance with such an assertion.<sup>83</sup> Since she should assert it given how she behaves one wants to treat her *as if* she had asserted it. In such a way one can agree or disagree with her even in the case of absent utterance. By relating to her in such a way we treat her as our interlocutor who hasn't yet said what she *should* say. Her actions speak louder than words, so to speak, and merit a response that would be relevant to an assertion. This is a case of absent utterance where the need to express one's engaged response makes explicit why the vicarious assertion is reasonable and why it motivates the introduction of the new term, "belief", to mark its difference from ordinary assertions.

My suggestion, then, is that the ascription of a belief is meant to bring out the fact that the assertion on someone else's behalf is appropriate when the other person ought to be ready, given her behavioural (and perhaps other) dispositions, to endorse such a vicarious assertion. This is appropriate even if she hasn't made or isn't disposed to perform that speech act. For instance, imagine someone who doesn't say that she thinks that people of different races deserve different treatment but treats them in discriminatory ways (Schwitzgebel 2010: 532). The attribution of a racist belief to her serves as a means for disagreeing with her, even though she hasn't made any corresponding assertion, and to bring her into the discursive space where it is hopefully possible to argue with her as if she had asserted it

Why have I stressed that the other *ought* to be ready to endorse the vicarious assertion? The motivation here is to respect the fact that our ordinary practice of belief attribution doesn't treat agents as fully authoritative over the convictions they actually have. There are cases when an assertion may be endorsed by a person but she can nevertheless be taken as believing something else or even contradictory to the original assertion. I have in mind here those occasions when people sincerely claim one thing but their actions speak against it. Such was the

---

<sup>81</sup> Believing something can be accompanied by a feeling of conviction but I don't think that this feeling is the essential characteristic of belief.

<sup>82</sup> One can also express this in the terms used by Ronald de Sousa: while desires have both semantic and emotional satisfaction conditions, beliefs can only be semantically satisfied i.e. are true or not (de Sousa 1998: 123).

<sup>83</sup> Such a way of behaving should be describable without any reference to beliefs in order to avoid circularity. The concept can't be already in use for it to be introduced. For a proposal how to specify ascription-relevant dispositions without reference to mental states, see Stout (2006).

case of the implicit racist, for instance. The belief attribution, then, isn't limited to situations wherein the other person endorses the vicarious assertion. It also extends to the situations wherein, given her overall behaviour, we think that she should agree.

As another example consider a situation, imagined by Dennett, wherein an art critic gives praise to his son's average paintings. Does he believe, Dennett asks, that his son's paintings are average and does he simply want to promote his son's career or is he so blinded by his love for his son that he really believes that these are high-quality paintings (Dennett 1981: 39)? Dennett presents this example in order to argue that there are cases when the facts about brain-wiring might be relevant for deciding what the person believes. In the present context, though, the ascription of the first belief (namely, that his son's paintings are average) exemplifies a situation wherein the belief attribution makes explicit what the person should assert, considering his overall dispositional profile as an art critic. In fact, it is the determining role of behavioural dispositions and the fact that the person might not be disposed to assert that *p* that motivate the introduction of the concept of belief. Otherwise we could simply talk about actual and potential assertions. The considerations about conceptual needs, then, support the view of belief attributions which stresses the wider dispositional profile in determining whether it's appropriate to ascribe a belief or not. This doesn't mean that in Dennett's example the attribution of the first belief is not contestable but it illustrates how beliefs are something over and above assertions.

As with our account of desire attributions the present proposal about the concept of belief has an advantage over Gauker's original communicative conception because it gives the concept an explanatory role. It articulates the conditions under which vicarious assertions in the case of absent utterance are appropriate but it avoids postulationism. The dispositional profiles that ground belief attributions don't need to imply anything about causally efficacious inner states. Nevertheless, the belief attribution itself has a genuine explanatory role to play. It doesn't merely state the vicarious assertion; it explains why it is appropriate to assert. Such an explanation makes intelligible our engaged responses to the agent by situating her behaviour in a wider context which includes more than her dispositions to assert. This is somewhat similar to Julia Tanney's account of reason-explanations according to which the latter don't pick out causes behind people's actions but rather give insight into wider circumstances which the action in question belongs to. Such an explanation gives additional information about the situation of the agent in order to make her intelligible to us (Tanney 1995; 2009).<sup>84</sup> In the present case, this kind of intelligibility amounts to the possibility of agreement or disagreement with the agent. As in Tanney's account, the belief attributions are meant to convey

---

<sup>84</sup> Since Tanney's analysis concerned action-explanation, the link between her and my account isn't that direct. Nevertheless the idea that belief-ascription don't mark causal, but contextual explanations, is something that we share.

further information about the agent than that which concerns her explicit assertions or even her dispositions to assert.

This doesn't mean, though, that the explanatory role of the concept of belief is limited to explaining why absent assertions on the behalf of others can be appropriate. The fact that belief attributions pick out the dispositional profiles of the attributees can be exploited in various situations. For instance, the attribution can identify the real reason why the agent acted, even if she asserts something different. In other cases the belief attribution simply gathers together patterns in the behaviour of the agent. But this can still be taken as a form of explanation because it brings forth the overall biographical background of the person to which that person's actions belong. What's more, one can rely on this attribution in future situations when someone asks about the person in question. The dispositional profile that grounds the belief attribution can involve various dispositions that have different manifestations in different conditions. Because of this, it allows the ascribers to make generalizations on the basis of the ascriptions which then can be appealed to when explaining the behaviour of the attributee (Schwitzgebel 2002: 266).

Also, the present explanation of the concept of belief can articulate the wider significance of belief attributions just as the explanation of the concept of desire did. Most notably, the extension of engaged responses of agreement/disagreement to the cases of absent utterance enables agents to subject each other to rational criticism when the person who is criticized hasn't explicitly said anything. Her behaviour indicates that she should assert that  $p$  because she acts as if  $p$  were true and if we want to criticize her actions we can also treat her as if she had asserted that  $p$ . The criticism isn't solely in the form "You shouldn't do it" anymore but can also appeal to the falsity of a person's belief as it is expressed in her behaviour. From the other side, the attribution of a belief, if one agrees with it, can be relied on to express reason-based support for the way someone has acted although she hasn't explicitly said anything.

There is still a gap in the current account. It doesn't say enough about the conditions under which vicarious assertions in the form of belief attributions are appropriate. This is basically what the problem of grounds was all about. It was the problem of grounds that motivated one to seek amendments to the original communicative conception. In the case of desires the gap was bridged by connecting the vicarious requests with the hedonic profile of those who were being interpreted so that desire-ascriptions would have a firm grounding. But in the case of belief attributions the appeal to behavioural dispositions is more vague because it doesn't pick out anything specific. It seems that such dispositions often underdetermine which belief attributions are appropriate.

Let's take, for instance, the well-known example that has been used as an intuition-pump in research on the theory of mind. It is about a man who runs towards a train that is beginning to leave the station. The explanation to this is couched in terms of a belief that there is a train that is just departing (Scholl & Leslie 1999: 131). Or take the much discussed case when a dog chases a cat and, after the latter has climbed on a tree, stays under it and barks. This seems

to allow explanation of its behaviour by attributing a belief that the cat is up there to the dog (Malcolm 1977, Davidson 1982). These are cases when mental state attributions have their sources solely in the nonverbal behaviour of those being interpreted. However, it is also apparent that these examples don't seem to motivate a particular belief attribution. Instead there seem to be various attributions available that equally fit the behavioural evidence (Stich 1983: 104; Putnam 1992: 28).

Since the modified communicative conception presupposes that the attributers and attributees are able to utter and comprehend assertions, there seem to be some additional means to delimit the range of acceptable belief attributions. The modified communicative conception can also treat the example of attributing beliefs to nonhuman animals as irrelevant because animals don't engage in linguistic practice. In the case of human beings one can make sense of a determinate assertion on another's behalf if this can be inferred from the speech act she has already performed herself. To take a rather boring example, if a person asserts that  $p \ \& \ q$  we can usually assert on her behalf that  $p$ . We can also presumably make vicarious assertions by relying on nondemonstrative inference. Consider the case when someone asserts that Quine is the most important philosopher of the 20th century. Then we can also tentatively assert on her behalf that deconstruction as a philosophical method shouldn't be taken seriously. But what can we say about the cases of absent utterance when there are no actual speech acts available. Such was the example of the man running to catch the train which was about to leave the station. These are supposed to be the situations wherein the conceptual need for the belief attributions first reveals itself. But what justifies making a particular assertion on another's behalf?

The question is quite tricky. I admitted that behavioural dispositions determine what kind of assertions on others' behalf count as belief attributions. But it doesn't seem to be the case that simple behavioural cues from the other can be used as a ground for asserting on her behalf. Something further is needed in order to do that. For instance, if I see a man running towards a train this information in itself doesn't suffice for asserting on his behalf that the train is leaving. If a postulationist is right and beliefs can be defined as causal postulates then the attribution of attitudes on the basis of behavioural cues seems to be easy – one must only make an inference to the beliefs. Things aren't actually that simple, though, because it is also necessary for a postulationist to bring in additional assumptions so that the explanation could be arrived at. These assumptions would include theoretical information, for instance, and possibly knowledge about the social setting. But it is still possible to see how the explanation can be worked out on the postulationist account – by using behavioural information coupled with theoretical and social knowledge. The communicative conception dispenses with the causal inference which naturally leaves a certain gap that needs to be fulfilled. If the ascription isn't based on causal inference, what could be a sufficient condition for ascription? I have already modified the original communicative conception so that it isn't in tension with our explanatory concerns. But the danger here is that there

wouldn't be any explanatory work left for vicarious assertions when it comes to accounting for the process by which we attribute beliefs if we allow for theoretical information. This would mean that some crucial questions would fall out of the purview of the communicative conception, at least when the attribution of beliefs is the issue.

How, then, can we constrain belief ascriptions in a satisfactory manner? It may seem surprising but the answer is that the proponent of the communicative conception may have to embrace a form of simulationism (Gordon 1986). We can assert on behalf of the man that the train will be leaving by considering his situation and thinking of what we would assert in that context. If we are ready to assert that the train is leaving then the corresponding belief can also be attributed to another person. This kind of proposal can supply a satisfactory constraint for the attribution because it enables us to see why I should assert this proposition – it is reasonable to perform those speech acts in the situation that interests us. Otherwise, the question would remain: why should I make exactly those attributions and not different ones? I could in principle also say that the man running after the train believes that the train is beautiful. But this doesn't (usually) seem to be a reasonable attribution.

One may argue that making assertions on the behalf of others can be more automatic than that. We just produce them under relevant occasions without any deliberation on the reasons for making them. But this kind of suggestion still owes us an explanation of why the automatic response to the question about the man running turns out to be as it does. Adopting a simulationist strand into the communicative conception seems, then, explains how one can assert something on another's behalf without the reliance on any verbal cues.<sup>85</sup> This conception of belief attribution is reminiscent of Jane Heal's account of co-cognition which involves the idea that we attribute to other people beliefs by thinking through the same subject-matter (Heal 1998). The crucial elaboration here is that the process of co-cognition is about imagining what one would assert if one were in another's position.

One can maintain, then, that the idea of beliefs as vicarious assertions survives as an important component in the modified communicative conception. In addition to Gauker's original idea, the present account claims that the need to talk specifically about beliefs arises in communicative contexts wherein we need to extend our engaged responses to the cases of absent utterance. Besides the simulationist proposal about how the process of belief attribution might work and besides the concession that the belief attributions should be seen as explanatory, the present account doesn't steer that far away from Gauker's original proposal. But as with desires, one can't remain so deflationist – the attribution of beliefs also explains why it is appropriate to assert on the behalf of others.

---

<sup>85</sup> I don't think that simulation is always necessary for constraining interpretation. For instance, when verbal cues are in place, co-cognition isn't needed.

## 5.4 Comparisons

### 5.4.1 Desires and the hedonic theory

Before exploring a few possible objections to the current account I compare it with some approaches that have the similar spirit. The aim of the comparison is to locate the communicative conception among likeminded theories and to show that its concerns are not cut off from theirs.

The account of desire given here doesn't seem to have any obvious precedents. It still takes some inspiration from the hedonic (pleasure-based) theories of desire according to which it is not essential for having a desire to instantiate a disposition to act. Rather, according to such theories, the more important connection is between desires and feelings of pleasure and displeasure. If someone has a desire that  $p$  the content of her desire is presented to her as something pleasurable. Galen Strawson is a well-known proponent of the pleasure-based theory. His main motivation comes from a thought experiment in which he envisages an intelligent race of creatures, the Weather Watchers, who lack any dispositions to act yet nonetheless have desires for various states of affairs and various things that please them. This seems to indicate that the capacity to feel pleasure can enable persons to want things even if they are not able to act (Strawson 1994, Ch. 9).

The modified communicative conception of desires and the pleasure-based theory differ in their overall aims. The first is about the ascription of desires while the second pretends to give an account of what desires are. Nevertheless, the pleasure-based theory, as Strawson conceives it, is based on the intuitively believable facts about desire attributions. Because of that one can still draw comparisons between the two accounts. Both reject the view that desires should be seen to be constitutively tied to actions and this rejection comes from intuitive considerations. Because of this they are largely in agreement about the negative. As for the relation between desire and pleasure, the modified communicative conception is more cautious. It admits the idea that the ascription of desires plays an explanatory role in accounting for the appropriateness of vicarious requests because it assumes that the subject would feel pleasure when the request is fulfilled. But this doesn't entail that the ascriptions of desires *always* need to take into account the pleurability of the outcome. That doesn't make the explanatory connection between desire and pleasure superfluous. This connection still identifies the normal conditions under which the ascription of desire is appropriate. The modified communicative conception explained why a concept different to request was introduced. It doesn't exclude the possibility that the new concept couldn't be ascribed in cases when that explanatory condition doesn't hold. The condition for possessing the concept of desire is that the subject can associate the fulfilment of desire with positive hedonic states. It is not a necessary condition for the success of a particular desire

attribution that desire be associated with pleasure. That being said, the pleasure-based theory of desire is still the most similar to my account that I could find.<sup>86</sup>

#### 5.4.2 Beliefs and deontic scorekeeping

We will now move on to the belief attributions. Some similarities can be found between Robert Brandom's account and mine. In Brandom's deontic scorekeeping model of communication, the ascription of propositional attitudes is meant to distinguish between different inferential commitments that agents attribute to others and undertake themselves. Before drawing some comparisons I need to devote some space to introducing his theory (which is very elaborate). It revolves around the idea that in participating in the communicative game the players attribute inferential commitments and entitlements to those commitments to one another. They also undertake such commitments themselves which they often need to justify to others to show that they are actually entitled to those commitments (Brandom 1994: Ch. 8). According to Brandom, then, linguistic communication is a normative enterprise where every move is an attempt to make a difference to one's normative status. So how does belief attribution enter the picture? The concepts of propositional attitudes in Brandom's account are in general meant to articulate the perspectivity of discourse. This means that attitude attributions make it possible for interlocutors to track the difference between commitments that they have and commitments that they attribute to others (Brandom 1994: 503).

The participants in communicative practice are individuals and they don't always share the commitments that others have. To make those differences explicit, they can speak of each other in terms of different beliefs and intentions. Belief attributions explicate the difference between theoretical commitments that people have and intention attributions make explicit the difference between practical commitments. Since what we are interested in is the concept of belief we should make this idea a bit more precise. As in the communicative conception, in Brandom's account the practice of asserting (or claiming) is the place where the belief-talk originates. People wouldn't have the status of believers without this (Brandom 1994: 153). So it's the tension between the

---

<sup>86</sup> Sander Voerman (2009) has proposed that the method of checking whether we got our practical judgements right proceeds through affective evaluation of the effects that come about when these judgements are followed. For Voerman, affective states can disconfirm practical beliefs. This doesn't directly connect this view with mine, because, after all, one could state that the knowledge about affective profiles can only inform the attribution of practical beliefs, but not desires. But still, if we interpret Voerman's view so that it also applies to desires – and practical belief can be taken as a type of desire -- it would be almost identical with the account I've defended. What's more, Voerman also takes the affective profile to reveal what a person *really* wants, thus making the similarities with my approach even more salient. A crucial difference of Voerman's approach from mine is that he defines desires as attitudes towards action-types.



doxastic commitments carried by assertions that forces us to speak about ourselves and others as believers.

One property of belief ascriptions that Brandom takes to be relevant for articulating the tension in question is that they can be given both a *de re* and a *de dicto* reading. When we read the ascription *de re* we can substitute the co-referring terms in the content-clause without changing the truth value of the ascription. In the case of *de dicto* reading substitutions may change the ascription's truth value. For Brandom the important aspect of that distinction is that ascriptions that are read *de re* express what we as attributers are talking and thinking about while *de dicto* ascriptions specify how the thought's object is being thought about by the attributee (Brandom 1994: 503). By ascribing a belief to another person the attributer can articulate the perspectival difference between the two of them. For instance, let's take a case wherein subject A ascribes to person B the belief that Cicero was a famous Roman orator. On the *de re* reading "B believes of Cicero that he was a famous Roman orator" A can substitute "Cicero" with "Tully" in the content-clause. On the *de dicto* reading "B believes that Cicero was a famous Roman orator" this can't be done. In the first case, A undertakes the substitutional inferential commitments of the term "Cicero". In this way she makes explicit her way of representing Cicero. In the second case, she only attributes the commitment to B and thus expresses B's perspective. Belief ascriptions, then, can be used to articulate the commitments of others in relation to one's own.

What is the wider pragmatic significance of belief-talk (and also desire-talk) in Brandom's account? It seems that, for Brandom, the difference in propositional attitudes of agents is something that gives communication its function in the first place. As he says: "if two interlocutors did (per impossible) have exactly the same beliefs and desires, communication would be superfluous" (Brandom 1994: 510). Without the difference between commitments that the agents undertake and attribute to each other, there would be no point in making any communicative moves. This may sound a bit excessive because communication doesn't seem to be only about challenging each other's inferential commitments. But when we acknowledge that Brandom includes also so-called language-exit moves (those that pertain to action) and language-entry moves (those that pertain to perception) among inferential commitments, his conception of communication and propositional attitudes professes to be comprehensive.

Inferential commitments, according to Brandom, have normative significance because they imply that the agent should do what she is committed to. My account locates belief attributions in a situation wherein the ascriber should assert something given her dispositional profile. Because of that, there seems to be certain parallels between Brandom's account and mine. Also in the modified communicative conception belief attribution makes explicit the doxastic commitments that the other has given her behaviour. Brandom's theory could perhaps even be interpreted as a more elaborate account of the communicative conception. Although he does not define beliefs in terms of vicarious assertions, Brandom's analysis of the *de re* and *de dicto* readings of belief attributions and

his more general description of the way in which the attributer can distinguish between one's own commitments and the commitments of the addressee don't contradict the communicative conception. It might rather give a more substantial account of how asserting on other's behalf is structured.

But there are at least two advantages that the modified communicative conception has over the deontic scorekeeping model. The first thing to note is that the picture given by the communicative conception seems to be a bit simpler. It requires less conceptual baggage than the deontic scorekeeping model (i.e. the whole commitment-entitlement structure). But what's more important is that the modified communicative conception has the explanatory dimension which Brandom's account seems to lack. The original communicative conception had a problem with specifying the conditions under which it was appropriate to make vicarious assertions. An analogous information problem may arise for Brandom's theory. One could ask when it is appropriate, especially in the case of absent utterance, to attribute different inferential commitments to other people. This seems to be an explanatory question. The modified communicative conception offered an answer to the problem of the grounds for attitude attributions while Brandom's account hasn't done that. In this sense the first has an advantage over the second.

This brings us quite smoothly to the second advantage. In the case of the original communicative conception, it was possible to doubt whether that conception was able to articulate the difference between (vicarious) assertions and beliefs. By modifying it, a more substantial difference was established which was largely related to the explanatory role of belief attributions. Brandom's theory of belief attributions may face the same initial problem. It isn't clear why the difference in inferential commitments and the *de re/de dicto* distinction couldn't be articulated solely in terms of assertions. After all, assertions can also be read both *de dicto* and *de re*, and Brandom himself uses assertions (or claims) to exemplify that distinction (Brandom 1994: 507). Also by simply ascribing an assertion to someone, it is possible to make the distinction between one's own inferential commitments and those that are attributed to the other. My modified communicative conception grappled with that problem and made the difference between assertions and beliefs more explicit. For that reason, Brandom's conception has a lacuna which the modified communicative conception doesn't have.<sup>87</sup>

All in all, the modified communicative conception has some similarities to other accounts, namely, the hedonic theory of desire and the deontic scorekeeping model of belief. Nonetheless, it is irreducible to them. I will now move on to consider the most salient possible objections that might be raised against my account.

---

<sup>87</sup> As for desires, Brandom's account is quite different from mine because he overlooks the connection between desires and hedonic states. He also seems to take the attribution of a desire to consist only in the acknowledgement of a type of practical commitment to an action. (Brandom 2000: 90).

## 5.5 Problems

### 5.5.1 Language first?

After the genealogical account of beliefs and desires has been established one may still have a nagging doubt that explaining attitude attributions in terms of linguistic acts puts the cart before the horse. At least in the light of empirical research, the claim that prelinguistic creatures are unable to ascribe propositional attitudes is less than certain. One can point towards the implicit false-belief tasks in developmental psychology which are meant to probe prelinguistic infants' putative ability to ascribe beliefs. The results of that research are under heated debate but for many authors they demonstrate that infants have such an ability (Baillargeon, Scott & He 2010; Kovács et al. 2010; Träuble, Marinović & Pauen 2010). This is a wide-reaching topic and the analysis of the developmental data would require very lengthy treatment. Here the right response is to stress the point that the present approach seeks to understand why reflective, adult subjects would need to ascribe beliefs and desires to each other and also recognize that need themselves. Having a language seems to be a precondition for asking that question in the first place.<sup>88</sup> This isn't meant to disparage the research in developmental psychology. I am just stressing that such research has different explanatory aims from those of the present approach.

This strict distinction between the two explanatory projects might raise a question if it amounts to the claim that adults, at least in their reflective moments, have different concepts of belief and desire than do children. Wouldn't they be thinking (and talking) about different kinds of attitude if this were so? This might raise the concern that the seeming univocality of those concepts gets lost. Why think that there's a strong conceptual discrepancy between the concepts that adults and children have? This is a rather intriguing problem and I will deal with the conceptual discrepancy in the next section because this topic does not strictly depend upon the relation between linguistic understanding and the attribution of propositional attitudes. But to anticipate a bit, research in developmental psychology contains results which support substantial conceptual differences between children and adults.<sup>89</sup> It shouldn't be dismissed at the outset.

---

<sup>88</sup> In that respect the current approach can perhaps be compared to Stanley Cavell's project of defending ordinary language philosophy as contributing to our self-knowledge by inviting us to projectively imagine what we should we say in imagined contexts (Cavell 1979: 146). The comparison of the current approach and Cavell's would require further analysis which unfortunately doesn't fall under the purview of this work.

<sup>89</sup> Most noticeable in that regard is Susan Carey's work on children's understanding of the concepts of number and matter, compared to adults (Carey 2009).

### 5.5.2 Right kind of attitudes?

I have now acknowledged the possibility that the modified communicative conception may account for specific concepts which may not be the object of interest for developmental psychologists, for example. Conceiving of the concepts of belief and desire as devices for explaining certain communicative moves might have the result that these concepts are different from those that young children have. I have responded to this concern. But the worries don't stop there. It may be also possible that the concepts of belief and desire that are explained here differ from those that are employed in certain areas of the philosophy of mind and philosophy of action. Because the present approach is basically constructionist – the question has been when it would be reasonable to introduce a new concept – it is possible to doubt that the constructed concept differs from the one we actually have. This wouldn't shake the communicative conception completely but might cut it off from the more usual philosophical endeavours.

It is interesting to notice that in the philosophy of mind *belief* and *desire* aren't always seen as univocal terms. It has been claimed that they don't refer to a single kind of attitude, thus expressing different concepts.<sup>90</sup> There being different kinds of belief and different kinds of desire isn't then such an uncommon idea after all. The present question is this. If the terms that are employed in the communicative conception express concepts that are identifiable with one of those kinds, do they cover both of them (the luckiest result!) or do they form a separate kind? In what follows, I will consider the basic distinctions between the kinds of concepts that *belief* and *desire* are meant to express, or at least those that I take as basic. I will start with desires.

### 5.5.3 Different kinds of desire

In the case of the concept of desire, one doesn't have to search for long to find claims about the ambiguity of the corresponding term. For instance, G. F. Schueler distinguishes between two kinds of desires: pro-attitudes and proper or real desires.<sup>91</sup> Desires in the pro-attitude sense include all those attitudes that move people to act. This sense of desire is usually employed when one gives an account of practical reasoning (Davidson 1963). Real desires, on the other hand, form a specific psychological kind that is only one among many motivators of action. Having the first kind of attitude consists in the subject approving (in some minimal sense) of a certain state of affairs. It is implied by any rationalizing explanation of action. Unlike pro-attitudes, real desires don't arise in the organism in virtue of rational deliberation but are caused by an environ-

---

<sup>90</sup> I'll give the examples below.

<sup>91</sup> For similar distinctions, see Nagel's (1970: 29) distinction between motivated and unmotivated desires and Alvarez's (2010: 53) distinction between rational and nonrational desires..

mental impulse (Schueler 1995: 36). In the case of pro-attitudes one can't ask if a person has done something intentionally, although one didn't have a desire (pro-attitude) to do that. The relevant pro-attitude is implicated by the intentional action in question. A real desire to do that need not be implicated in such a way (Ibid. 29). One could say, for instance, that a boy left the house because he was told that he should go to school, believed that he should do what his parents tell him to and wanted<sub>1</sub> (had a pro-attitude) to go to school but he didn't want<sub>2</sub> (didn't have a real desire) to go to school. By his reasoning he approves of going to school but he doesn't really want to do that. He only has a pro-attitude towards the action.

Schueler's distinction is somewhat difficult to follow because real desires seem to be defined largely negatively as states that are not merely pro-attitudes. The paradigmatic real desires are for him simple cravings like hunger and thirst (Ibid. 34). But this seems too restrictive. One can't build an account of desires on such a limited set of needs. My impression is that the notion of proper desire seems to be "thicker" than the notion of a pro-attitude. The attribution of it has stronger psychological commitments. There may have to be some sort of psychological structure in place to have a real desire and there also might be a phenomenal character that pertains to that kind of desire. We could identify the relevant psychological structure and the phenomenal character with the hedonic state of the subject. This would make the notion of real desire resemble the concept given by the modified communicative conception.<sup>92</sup> In the example of the boy going to school the reason why we can say that he actually didn't want to go to school may be that there was no pleasure or positive feeling associated with such an action.

But what can the modified communicative conception say about pro-attitudes? Do they fall under a different concept than the one that the modified communicative conception aims at explaining? For Schueler pro-attitudes were necessarily implicated in the explanations of intentional action. Whenever X performed an action it was supposed to be explainable in terms of X wanting (having a pro-attitude) to achieve a certain state of affairs and believing that performing the action in question brings about that state of affairs. It seems that the modified communicative conception employs a thicker concept of desire and can't account for the concept which designates pro-attitudes which don't have such an inferential commitment. This is probably a surprising result. The original communicative conception seemed to be much more in line with the pro-attitude approach because it didn't say anything about the connection between desires and pleasure. But let's assume that the modified conception is to be preferred here. What does it have to say about pro-attitudes?

One possible response here would be to claim that the genealogical account wasn't meant to cover desires in the sense of pro-attitudes which are concocted by philosophers and aren't of much interest in the study of our commonsense practice. But I do think that pro-attitudes can be accommodated in the present

---

<sup>92</sup> This is a very loose interpretation of Schueler's position.

account if we consider what the communicative conception is about. As already noted in §4.4.1, desires' connection with pleasure doesn't require that the attribution of desires always associates a positive hedonic state with their fulfilment. The explanation was only meant to bring out the primary rationale for the introduction of that concept by articulating the conceptual connections it enters into in normal situations. Unlike the traditional conceptual analysis, the modified communicative conception doesn't determine the necessary and sufficient conditions for the concept's employment.<sup>93</sup> Minimally speaking, the attribution of a desire that  $p$  to  $X$  is simply a request that  $p$  on  $X$ 's behalf. The attribution can always be paraphrased into the corresponding vicarious command. This also applies to explanations of intentional action which appeal to the desires (i.e. pro-attitudes) of the actors.<sup>94</sup>

There is one obvious objection to the present attempt at accommodating pro-attitudes in the modified communicative conception. The complaint is that by allowing desire attributions that don't commit the attributer any assumption about the attributee's hedonic profile, my account collapses back into Gauker's original proposal. This proposal didn't make any assumptions about the connection between desire and pleasure. If the communicative conception is meant to cover both pro-attitudes and real desires, the original account seems to achieve that more smoothly.

I disagree. Aside from answering the question about the pragmatic significance of the concept of desire (which is already an important achievement) my modification of Gauker's account also explains why desire attributions are normally governed by the assumption that the subject should be subjectively satisfied when the desire is fulfilled. There is still something wrong with the desire-attribution when she isn't. Even in cases when only the pro-attitude is attributed and no explicit assumption is made about the subject's hedonic states, the attribution can still be implicitly governed by such an assumption. This becomes apparent in situations when the fulfilment of the desire and the subjective satisfaction don't match – the concept of desire doesn't seem to play its proper role in such a case. The assumption that the attributee should be subjectively satisfied with the desire's fulfilment can be explained by the modified communicative conception. It can't be explained by the original version of the communicative conception which didn't distinguish between commands and desires sharply enough. This is an intuitive point about what is assumed about desire attributions but the idea of a fulfilment of desire without a subjective satisfaction seems to be very alienating. If the concept of desire were *entirely* disconnected from the idea of such a satisfaction – something that the original communicative conception would allow – it loses its distinctive meaning.

---

<sup>93</sup> Similar point about knowledge was made by Edward Craig in his genealogy (Craig 1990: 14).

<sup>94</sup> I actually doubt that an explanation of an intentional action necessarily appeals to something like pro-attitudes – an acceptable explanation can perhaps be also accounted by the actor submitting to relevant norms - but I am not going to speculate about that here.

It seems, then, that the modified communicative conception is able to account both for real desires and pro-attitudes. That being said, I have to admit that I'm not fully convinced that admitting ambiguity in the term "desire" is avoidable. One can argue that the original communicative conception accounts for desires in the sense of pro-attitudes, while the modified version applies only to desires proper. A possibility of such a division of labour should also be kept in mind in the next section with regard to beliefs.

#### 5.5.4 Different kinds of belief

In the case of the concept of belief there are also distinctions that can be drawn which would make the concept ambiguous. For instance, one could distinguish between occurrent and dispositional beliefs and see these as different kinds of mental states. Following Frankish (2004), one could also map this distinction onto the difference between explicit and implicit or personal and subpersonal beliefs. I suspect that commonsense conceives of beliefs in the dispositional sense while the idea of an occurrent belief has its use in certain philosophical accounts. Having a belief is a state not an event (according to Williamson 2000: 35) and merely entertaining that  $p$  or even judging that  $p$  doesn't amount to believing that  $p$ . If someone talks about occurrent beliefs she can be taken to be speaking about something that shouldn't be equated with proper beliefs. That being said, these remarks don't suffice to preclude using the notion of occurrent belief for technical purposes in the philosophy of mind. But they give a reason in the current context to limit oneself to discussing beliefs in the dispositional or standing-state sense. The communicative conception of belief should account for them.

It does seem that the modified communicative conception that I've defended is able to explain the same facts that the theory of dispositional beliefs seeks to understand. It is well suited to account for the attribution of dispositional beliefs because those assertions on the behalf of others which constitute belief-ascriptions take into account more than the actual utterances of others. They are meant to pick out the behavioural dispositions of the attributees. This explains the need to talk about dispositional beliefs in a way that transcends what the attributee actually asserts. This, of course, doesn't mean that the communicative conception, even its modified version, can be equated with the dispositional theory. Rather, the communicative conception can account for the fact that the ascribers of belief take into consideration the dispositions of the attributees. But it doesn't claim that beliefs *are* complex behavioural dispositions as the dispositional account does (Gauker 1994: 275).

But it is also possible to distinguish between two kinds of dispositional belief. There is the idea of beliefs that relates them more closely to occurrent judgements, so that the attribution of a belief that  $p$  to  $X$  relies on  $X$ 's

disposition to consciously judge that  $p$  (Gendler 2008).<sup>95</sup> One of the most obvious routes to the attribution of a belief that  $p$  to a person does seem to be answering the question whether she tends to affirm  $p$ . A deeper reason for individuating beliefs in terms of explicit judgements is that in judging we are responsive to evidence and this would also explain the responsiveness of beliefs to reasons. Another way to employ the dispositional concept of belief is to consider the more inclusive dispositional profile of the person whose beliefs we are interested in. In that case we don't take explicit judgements to be the only or even the main determinant of having a belief. According to such a view, the explicit judgement that  $p$  can be rather insignificant and other kinds of behavioural dispositions are much more important when we individuate beliefs (Hunter 2011).

We can express the difference between these two positions in terms of the first-personal and third-personal approaches. The view that individuates beliefs in terms of explicit judgements stresses their first-person accessibility. The position according to which broader dispositional profiles are central, on the other hand, focuses on the role of belief attributions in action-explanations and is third-personal (Zimmerman 2007). I draw attention to this distinction because we get differences between attributions in these two cases. Consider (again) the implicit racist who claims that she believes in the equality of all races but who acts towards people from a different race in such a way that is degrading or disrespectful. According to the first-personal approach, we should take the assertions of a person at face value and attribute to her the belief that all races are equal - after all, he would consciously judge that all races are equal. He is consciously aware of what he judges. According to the third-personal approach, the attributions of belief rely on the facts about the attributees' behaviour. The implicit racist believes that all races are not equal. The facts about his behavioural dispositions need not be consciously accessible to him and other people might be in a better position to know about them. The two approaches, then, decide differently on whether we should attribute a belief. How should the modified communicative conception treat those cases?

I already hinted above that the communicative conception accommodates the third-personal standpoint and claims that it is proper to attribute to the implicit racist the belief that all races are not equal. The reason is that the modified communicative conception acknowledges that the attribution of a belief is appropriate when the attributee's behavioural dispositions indicate that she *should* endorse the relevant assertion. This is a result that we should want. The first-personal approach according to which beliefs are ascribed on the basis of mere dispositions to judge is controversial because one can imagine cases wherein people do believe that  $p$  but are not disposed to judge that  $p$ . For

---

<sup>95</sup> For Gendler, instantiating a mere behavioural disposition in relation to a state of affairs  $p$  which doesn't include the disposition to consciously judge that  $p$ , shouldn't be taken as a case of having a belief that  $p$  but rather an alief. I am not entirely sure that Gendler can be neatly associated with the view under consideration here but she surely dismisses the broader dispositional profile as being irrelevant in determining what a person believes.



instance, sometimes a person is too psychologically damaged to affirm the content which she believes (Cassam 2010: 83). One could try to elaborate on the first-person disposition to judge by stating *ceteris paribus* conditions. But it seems more reasonable to simply infer that there are behavioural dispositions that are also relevant in determining when the attribution is appropriate. We shouldn't limit evidence in favour of vicarious utterances to only the explicit judgements of the attributee. The third-personal approach is more in line with the modified communicative conception.<sup>96</sup>

Notice that the communicative conception doesn't equate beliefs with broad behavioural dispositions. Instead it sees the latter as forming a relevant attribution-base. Doubt may be raised whether it is a too wavering position because of that. In the dispositional account the identification of beliefs with dispositions also fully explains why belief attributions which rely on dispositions are true. In the case of the communicative conception it isn't totally transparent why attributions that rely on dispositions reveal the beliefs of others.

But the indeterminacy of belief attributions can actually be an advantage. It enables a certain amount of negotiability about belief attributions. The ambiguity of a concept need not be taken as something negative so that it has to be made more precise. The fact that the self- or other-interpretation isn't fully determined by the underlying dispositional profiles of agents enables us to maintain a kind of interpretive openness. The contestability of belief attributions need not be their weakness. Admitting that doesn't defeat or cancel the point that broader dispositional profiles *are* relevant for determining whether the attribution is appropriate. This is exactly what distinguishes belief ascriptions from mere vicarious assertions. The modified communicative conception, then, clearly fits better with the idea that beliefs are individuated in terms of behavioural dispositions, not merely dispositions to judge or assert.

## 5.6 Conclusion

In this chapter I criticized Gauker's communicative conception of beliefs and desires by challenging him to explain the difference between attitude ascriptions and speech acts on others' behalf. Although Gauker wanted to keep his commitments minimal he still had to explain the need for separate concepts, and my complaint was that he wasn't able to do that. To amend this problem I proposed a modified version of the communicative conception. According to it the attribution of beliefs and desires allows us to have engaged responses of agreement/disagreement and endorsement/disapproval to other people even in situations when they haven't performed the respective speech act. These attributions make intelligible why it is appropriate to assert and request on others' behalf in the case of absent utterance. They have to take into account

---

<sup>96</sup> What about first-person authority of belief attributions according to the communicative conception? This is a question for a separate chapter.

certain facts about the attributees, namely, their behavioural and hedonic dispositions. Nevertheless, the modified communicative conception still retains the central idea of Gauker's original account, namely, that belief and desire attributions involve assertions and requests on the behalf of others, thus making possible the engaged responses to the attributees. In the following chapter I will ask whether the modified communicative conception (*qua* an account of mental state attributions) is also an account of the nature of beliefs and desires.

## 6. THE COMMUNICATIVE CONCEPTION AND ONTOLOGY

### 6.1 Introduction

Analysing some area of vocabulary in terms of what it enables us to do and what needs it fulfils can certainly be a fruitful exercise. Another question is whether this kind of approach could also say something substantial about the phenomena which this vocabulary is supposed to refer to. Especially since Kripke and the postulation of *a posteriori* necessities, it is common to distinguish between the nature of a phenomenon and the way in which we speak and think about it. Our epistemic approach to some thing or property is mediated through our conception of what other properties relate to it. But it is possible that the essence of the thing or property in question is independent of the descriptions that we associate with it. The properties of being drinkable and transparent don't reveal the nature of water. The property of being yellow doesn't reveal the nature of gold and the property of having a minor interest in stamp-collecting doesn't reveal my nature.

Can a similar distinction be drawn in case of propositional attitudes? Do the facts about the ascription of beliefs and desires underdetermine the nature of those states? In a certain sense, the underdetermination applies trivially because the facts about attribution can also include the errors that the subjects make, for instance. This already suffices for concluding that the nature of mental states can't be simply read off from the facts about their ascription. A choice has to be made by asking which ascription-facts are relevant for understanding the ontology of mental states. This implication of *a posteriori* necessities goes even deeper. It might be argued that *no* set of ascription-facts can fix the facts about the nature of mental states. For instance, a robust causal theory of reference could claim that we can refer to beliefs and desires even when the attribution of mental states as a whole proceeds only through a contingent, and possibly false, understanding of those states. The idea that the nature of mental states is determined by the facts about their attribution seems to put the cart before the horse and bring us to a dubious form of anti-realism. According to this view the world needs to respond to our concepts, not the other way around. So at least a preliminary look at the issue may pull us towards the view that the nature of beliefs and desires can't be inferred from facts about their attribution. Let's call it the Ontology-without-Epistemology view (OwE for short)

OwE seems to be a radical view because it severs the link between what the folk think of attitudes and what these attitudes really are. The present chapter seeks to find a more moderate position on what ontological conclusions can be drawn from the version of the communicative conception of beliefs and desires I am defending. According to the modified communicative conception (MCC), the attribution of a desire that *p* to *X* can be identified with requesting *p* on *X*'s behalf, on the assumption that *p* would please *X*. The attribution of a belief that *p* to *X*, on the other hand, can be identified with asserting *p* on *X*'s behalf on the

assumption that X should assert that p herself. These claims were meant to bring out the practical significance of those concepts by articulating their application conditions. How ontological conclusions can be drawn from such application-conditions is an open question. OwE is only one possible (and negative) answer.<sup>97</sup> In this chapter I will tackle this question and investigate some alternatives to OwE.

The plan of this chapter is as follows. First, I will reject the view that we can (largely) dismiss the facts about the attribution of beliefs and desires when we are interested in their nature. Second, I will present two opposing views of the connection between attribution-facts and ontology. The first is conceptual analysis, exemplified by analytic functionalism. The second is the view that we can read off the nature of beliefs and desires from the use of the respective concepts. I will argue for the second approach and claim that it helps us articulate the minimal ontological commitments that the attributers of belief and desire have. In the second part of this chapter I apply this approach to the question, what does the modified communicative conception entail about the ontological status of beliefs and desires? I reject the (strong) interpretationist interpretation of MCC according to which interpretation-dependence follows from the communicative conception and settle for a dispositionalist and realist (non-interpretationist) view according to which the (true) attributions of beliefs and desires can be true and that they reveal interpretation-independent facts about behavioural and hedonic dispositions of persons.<sup>98</sup>

## 6.2 From epistemology to ontology

We started this chapter by sketching out OwE. This is the claim that our folk psychological practice of attributing beliefs and desires works on assumptions that might misconceive the nature of those states to a large extent, so that one can't infer anything about the latter from these assumptions. Nevertheless, we succeed in referring to those states and it is the task of the natural sciences to discover their nature.<sup>99</sup> This view can be motivated by the causal theory of reference which maintains that a term's reference is determined by its original introduction and the causal chain that connects the use of the term with its original introduction. According to this view, a term can maintain its reference even if various descriptions that are associated with it turn out to be false. A

---

<sup>97</sup> Gauker himself takes the communicative conception to be both about the attribution and about the ontology because he adopts an interpretationist position (Gauker 1994). One of the purposes of this chapter is also to investigate whether Gauker is justified to do that.

<sup>98</sup> What realism about an area of discourse actually amounts to is a contested issue, of course. Here I simply follow what I take to be an intuitive understanding.

<sup>99</sup> A more accurate expression for OwE would probably be Ontology-without-Folk-Epistemology view because it is hard to believe that no kind of epistemic access to a property is relevant in trying to give an account of its nature.

good example of this outlook can be found from the rather famous passage by William Lycan:

I am entirely willing to give up fairly large chunks of our commonsensical . . . theory of belief or of desire (or of almost anything else) and decide that we were just wrong about a lot of things, without drawing the inference that we are no longer talking about belief or desire. (Lycan 1988: 31–32)<sup>100</sup>

This kind of outlook can be found in the work of various authors. For instance, Robert Cummins (2010: 68) argues against anti-individualism about belief content by claiming that if we want to understand what beliefs are we should look into scientific psychology. This means that many commonsense assumptions about belief can be dismissed. Or take Timothy Schroeder who tries to give an account of the nature of desires that doesn't need to respect every aspect of our folk conception and relies on neuropsychology instead (Schroeder 2004).<sup>101</sup>

For instance, if the communicative conception assumes that attributing beliefs and desires consists of vicarious utterances that can be made in specific situations, a proponent of Owe could perhaps even accept such an outlook. Nevertheless, she might argue that beliefs and desires can still be identified with some brain structures and assume what Gauker calls "postulationism" at the ontological level. I do think, though, that this would be a highly unstable position to hold. After all, the communicative conception was originally meant to counter the postulationist view that identified beliefs and desires with causally efficacious inner states. Conceiving of attitude-attributions as vicarious utterances and then claiming that the attitudes should be taken as inner states, the nature of which is discoverable by neuroscience, isn't tenable. We would need to introduce further assumptions that would explain why this kind of move is necessary.

This still leaves open the issue of whether Owe as such is tenable but in the present context, where the communicative conception of mental states is on focus, it is a secondary question. In any case, Owe expresses an entirely different sentiment to that expressed by my project as a whole. I assume that the nature of beliefs and desires should be recognizable by us so that we can take them as *our* attitudes. An account of the nature of attitudes that is cut off from our pragmatic concerns simply doesn't gain a foothold here. The question we face now is how exactly the folk psychological practice reveals our ontological commitments.

Nonetheless, I will briefly consider Owe for completeness's sake and ask whether it is a tenable position. If it is, the communicative conception faces a larger set of possible approaches to the nature of beliefs and desires. To put it bluntly, I reject Owe but I do that not on the basis of arguing for or against a

---

<sup>100</sup> Lycan's quote functions only as an exemplification, I don't intend to offer an adequate interpretation of his general approach.

<sup>101</sup> It isn't clear whether Cummins could claim that all folk assumptions could be ontologically incorrect. Schroeder certainly doesn't claim that.

specific theory of reference. I argue that Lycan's intuition can be dismissed because it's quite vague and difficult to explicate. The vagueness comes from the fact that Lycan doesn't make a full-blown claim that the folk conception is entirely irrelevant to sorting out commitments about the nature of belief. He talks of the willingness to give up only "fairly large chunks" of our folk conception of attitudes but not of giving it up as a whole. A question can be posed: what are the remaining "chunks" that Lycan is not willing to give up? Acknowledging that there are assumptions that can't be given up would take him too close to philosophers who admit that there are connections into which belief and desire enter which are definitive of the functional role of these attitudes. The commonsense conception would still largely determine what these attitudes can be.

Another problem with Lycan's intuition becomes apparent when we consider how it would look to give up the "chunks". Could we give up the idea that beliefs and desires are semantically evaluable when future neuroscience dispenses with that idea? Or could we give up the idea that beliefs and desires rationalize people's actions? Or the idea that we can respond to our attributions of beliefs and desires by agreeing/disagreeing with the former and satisfying/frustrating the latter? Note that by presenting these questions I am not implying that these possibilities are excluded. After all, philosophers already argue what belongs to the nature of those attitudes and what doesn't all the time (in fact, I am going to propose my own way to dismiss certain intuitively acceptable assumptions about beliefs and desires later in the analysis). The reason why these questions pose a difficulty here is that without being explicit about what the rejectable "chunks" are, the intuition remains too unarticulated. One should have at least some idea of how the progress of science would gradually challenge those assumptions so that some of them would remain accepted and others wouldn't.

Owing to these considerations, I assume that OWE is mistaken, if not confused. Our practice of attribution remains relevant to our project of saying something about the putative nature of beliefs and desires.<sup>102</sup> The question we face now is how exactly this practice reveals our ontological commitments. In what follows, I will consider two possible approaches to that issue: the aforementioned functionalist approach, inspired by Lewis's work and the approach that I, for the lack of a better word, call the "pragmatic view". I will eventually defend the latter but first I need to articulate the difference between the two.

---

<sup>102</sup> One should tread the ground carefully and stress the notion "putative" here because by admitting the importance of folk assumptions we can't say that the attributions of beliefs and desires can be true. This would require responding to eliminativism which I haven't done yet.

## 6.3 Two ways to understand folk commitments

### 6.3.1 Lewis/Jackson view

The first way to understand what our folk ontological commitments amount to is closely related to the programme of conceptual analysis. The latter need not be understood in terms of providing a set of necessary and sufficient conditions for the concept-application. It can also be taken to probe our intuitions about the concept in order to reveal our tacit folk theory. This theory is supposed to be the source of those intuitions (Lewis 1972; Jackson 1998: 32). A further move is to compare the folk theory with the one provided by scientific psychology and to see what the latter is committed to. If our folk theory's commitments overlap with those of scientific psychology then we can claim that the posits that are given by folk psychology are functionally realized by those states that are acknowledged by mature scientific psychology. In that case the ontological commitments of folk psychology could be vindicated (Jackson & Pettit 1993). According to this picture, then, the move from the attribution of beliefs and desires to their ontology comes in the form of an explication of the folk theory behind our everyday attributions of those attitudes and then giving an account of the ontological commitments of such a theory. This way of approaching the issue makes the everyday concepts vulnerable to scientific disconfirmation but also opens up the possibility of vindicating folk psychology by scientific means. It also assumes that folk concepts need to be naturalizable in order to be ontologically respectable entities.

How can we pin down the ontological commitments of a folk theory? In order for the commitments of scientific psychology and folk psychology to be comparable, the right method for identifying them should proceed in a similar way. I assume that the most standard way of doing it is to rely on the Quinean, or what is taken as Quinean, understanding of ontological commitments. According to Quine, in order to determine what exists we should consider our best scientific theories and see what has to exist for those theories to be true. By translating those theories into an appropriate formal language we can say what exists by looking at what realizes the values that variables quantify over (Quine 1948). This programme also seems to be applicable to folk theories in which case we can ask what must exist for them to be true.<sup>103</sup> Because of this connection the Quinian method of specifying ontological commitments can also be applied to folk psychology. One should simply look for what has to exist in order for the folk psychological theory to be true. This kind of approach can be seen as an attempt to compare those commitments with those of scientific psychology in order to see whether they are compatible. The point here is not that the Quinian approach to ontology and the Lewis/Jackson position are the same but that they share some important similarities. The second seems to be

---

<sup>103</sup> It is completely independent question whether Quine would have taken the conceptual analysis even seriously, considering that he wasn't interested in vindicating pre-scientific conceptions of the mind and world.

indebted to the first, at least in its general outlook when it comes to solving ontological issues.<sup>104</sup> Although Quine is taken as an arch-enemy of conceptual analysis, the *method* of pinning down ontological commitments is common to them both. Existence is still equated with being a value of a variable.

So there are two moments in the aforementioned programme that are crucial. First, it assumes that intuitive judgements about a concept by the folk can reveal the implicit theory that this concept implies. Second, such an implicit theory can then be investigated in order to see what entities, defined solely by their functional roles, it is committed to; namely, those that make the sentences of the theory true. But there is an additional move: in order to know whether those entities really exist we need to compare the folk theory with our best scientific theory. By doing that we can see whether the functional roles that the folk theory is committed to can be realized by what scientific psychology accepts. This final move doesn't belong to the task of conceptual analysis. The hope is simply that the functional roles, specified by the folk theory, can be realized in some way (Jackson & Pettit 1993). Following O'Leary-Hawthorne and Price, let's call this kind of approach to investigating folk commitments, the "Canberra plan", on account of the movement originated in the Australian National University in Canberra (O'Leary-Hawthorne & Price 1996).

As already hinted at in the introduction, it seems that the communicative conception of beliefs and desires that I have envisioned can't fit very neatly into the Jackson/Lewis programme. The main reason is that the modified communicative conception (MCC) didn't aim at providing a conceptual analysis which could reveal a theory in which beliefs and desires are embedded. It only claimed something about the way in which the attribution of those attitudes functions. Also, one could argue that, unlike MCC, the folk are intuitively committed to the view that beliefs and desires are causally efficacious inner states that are ascribed in order to predict behaviour. That is a position that the communicative conception opposes but seems to be a commonplace view in the literature about folk intuitions (Malle 2004; Malle & Knobe 1997). It seems that the communicative conception isn't well suited to the programme in question. Of course, that may be the fault of the communicative conception, not the programme itself. The fault here is a lack of explicitness about the folk theory that commonsense psychology involves.

In fact, Steffen Borge has presented the empirical studies of folk intuitions as direct evidence against Gauker's theory of communication but also against his conception of beliefs and desires (Borge 2013). He points out that the folk view of beliefs and desires gives them an explanatory priority to speech acts. This is something that the communicative conception seems to deny. I don't think that Borge's challenge necessarily damages the communicative conception or at least not the modified one that I have defended. The communicative conception,

---

<sup>104</sup> We get the taste of the Quinean spirit from the Lewis's approach when the latter claims that "we have a reason to think a good theory true; and if a theory is true, then whatever exists according to that theory really *does* exist." (Lewis 1972: 254)



as I envision it, just isn't a theory about the psychological processes that underlie folk psychological attributions. It isn't concerned with the intuitions that are supposedly an expression of such processes. Instead, it is an approach to the concepts of belief and desire that attempts to construe them in a functionally explicit manner in order to see the usefulness of those concepts. Another reason not to be struck by Borge's challenge is that Borge's arguments are intended mostly against Gauker's view of linguistic communication which doesn't give any substantial explanatory role to mental state attributions. The modified communicative conception of beliefs and desires, on the other hand, can acknowledge the full importance of belief and desire attributions in linguistic practice. But even if Borge's challenge doesn't directly affect the present version of the communicative conception, it still indicates that the latter is ill-suited for the Canberra plan which relies on folk intuitions. Is there a way to understand folk commitments that doesn't have the implication that people's intuitions reveal a tacit theory?

### **6.3.2 The pragmatic view**

Hutto proposes that when we are interested in folk commitments we should look at what people actually do when they attribute mental states (Hutto 2013). This method is an alternative to the method of probing the intuitive judgements about the applicability of those concepts in rather far-fetched scenarios as conceptual analysts usually do. It thus differs quite crucially from the method of conceptual analysis, envisioned above, because it doesn't aim at representing folk theories of the concepts of belief and desire. It focuses on the performance of competent attributers instead. Let's call it "the pragmatic view" because it focuses not so much on the semantics of propositional attitude attributions but the use of those attributions in everyday practice. One can find a similar sentiment also in Huw Price's pragmatist approach to different discourses (Price 1988; 1992). It certainly seems that the pragmatic view is better suited to be the methodological background of the modified communicative conception.

Unlike the Canberra plan, though, it isn't clear that the pragmatic view can even lay down any concrete proposal about how to discover folk psychology's ontological commitments. It doesn't presume that people employ a tacitly represented folk theory when attributing beliefs and desires. But without such a presumption it lacks the standard tools for pinning down ontological commitments. These involve perspicuously representing the folk theory to see what the theory must quantify over in order to be true. The pragmatic view seems to be ontologically too shallow and no substantial ontological conclusions about the existence or non-existence of beliefs and desires can be drawn from it.

Of course, triviality of ontological questions isn't entirely alien to contemporary philosophers. Huw Price, for instance, advocates metaphysical pluralism, according to which an affirmative answer to existence-questions is relative to a linguistic framework (Price 1992). Assuming that different types of linguistic discourses have different functions, the question of whether entities talked about

in a discourse exist should be solved within that discourse. So, simply put, mathematical entities can be said to exist within the discourse of mathematicians, moral entities exist within the discourse of morals, semantic entities are ontologically respectable within semantic talk, etc. It seems, then, that for this kind of view there can only be a purely pragmatic justification of ontological commitments. Since pragmatic concerns don't remain constant across different discourses, we can be ontological pluralists. In fact, questions about the ontology of mental states turn out to be so trivial that the whole matter turns on an analysis of the functions of different vocabularies (Price 2007). Price's theory seems to assume a quietist approach to matters of ontology because different discourses don't face any challenge from other discourses due to their functional distinctiveness.<sup>105</sup>

Stephen Stich is another author whose later views represent a pragmatist position. He is definitely quite promiscuous in his making of ontological commitments. This becomes apparent when he claims that there are innumerable properties such as 'couch', 'war', 'throws', 'fixes' and 'crushes'. These are not naturalizable by means of identification with some respectable predicate postulated in natural science. This doesn't mean that there are no couches or wars or that nobody ever throws, fixes or crushes anything (Stich 1996: 176). For something to be ontologically respectable, one doesn't even need to explain how it supervenes on the intrinsic properties of a thing. Stich uses the property of being an original Picasso painting as an example of a property that is clearly instantiated but can't supervene on the intrinsic properties of the canvass and paint (Ibid. 181). Stich also rejects weaker versions of the supervenience thesis such as those which identify the supervenience base with physical properties (Ibid. 184) or with nonintentional properties (Ibid. 185). It seems, then, that Stich takes ontological questions to be trivial. As long as the concept in question has a pragmatically justified use in human practice we can allow for the existence of the thing or the property that this concept represents.

To be fair to Stich, he does consider one way of evaluating the ontological status of entities that a discourse posits. His pragmatist proposal for seeing whether the entities of a discourse can be vindicated, which he is ultimately skeptical about, is that we should look at the history of science and analyse the cases when some posits have been rejected (witches, phlogiston, etc.) and when they haven't (mountains, tables, etc.) and try to formulate principles that govern those choices (Stich 1996: 64). After having done that, it is supposedly possible to apply those principles to the problematic discourse in question. But ultimately Stich rejects this possibility and sticks to his pragmatist guns by pointing out that picking out the right principle is an interest-relative matter and it is dubious whether different theorists share the same goals (Ibid. 66). The

---

<sup>105</sup> So far I have only employed the Pricean approach in order to do the functional analysis of mental state concepts, not to do ontology, and it seems that the latter doesn't give any insight how to answer ontological questions, rather, it suggests us to dismiss them. At least I think this should be taken as a default interpretation of his particular position.

choices of theories that have been made during the history of science have been driven by diverse factors (political interests among others) (Ibid. 67). Because of these considerations, Stich is doubtful that philosophers would be able to provide any clear normative advice about how to make ontological decisions.

The only constraint on ontology that Stich really accepts is that respectable entities need to have a role in some successful scientific theory (Stich 1996: 199). Funnily enough, this may already be too strong a constraint in the present context. We are here interested in the commonsense concepts of belief and desire. A claim that they are ontologically respectable only when they can be accommodated into some successful scientific theory is already too restrictive.

A third place to look for the pragmatic approach is in what many call Ordinary Language philosophy. Its proponents are also interested in accounting for the use of our everyday concepts. It isn't surprising that Huw Price is ready to associate his pragmatic approach both with Ryle and the later Wittgenstein. But again, also in the case of the latter authors, it is hard to get any suggestions about how to investigate ontological issues. After all, they were quite hostile to the idea of substantial metaphysics as such. The same applies to Austin. In the case of all these authors the stress is on the use of concepts and what they enable us to do in the context of human needs and interests (Baz 2010: 48).

Although it is hard to pin down the commitments of the pragmatic view, it still can be said that there are some commitments that it does *not* make. For instance, it is plausible to assume that the pragmatic view doesn't commit ordinary folk to the view that beliefs and desires are causally efficacious inner states or that those states need to be identifiable with or realizable by some physical states. As long as there are needs that these concepts can fulfil and criteria by which to apply them, they can remain in use. This doesn't depend on whether their application is responsive to some "deeper" matters of fact. Note that this relation between the pragmatic view and a denial of strong commitments isn't a relation of entailment, though. Since the pragmatic approach is loosely defined in terms of concept-use, it is more like an agnosticism about ontological commitments. It refuses to stake a stand on which kinds of entities the folk are committed to.

The pragmatic view also seems to involve quietism about folk concepts because changing or dismissing concepts on ontological grounds would likely have unseen consequences for the satisfaction of our pragmatic needs. To use an analogy, a proponent of the view in question can be compared with an individual who is politically conservative. She has no opinion on which principles of government are the right ones and she thinks that it is better to retain the current political system because nothing could guarantee that changing it wouldn't make matters worse. Quietism isn't the same as agnosticism – one doesn't entail another – but the pragmatic approach seems to be drawn to both.

So there are two opposing or at least divergent ways of approaching the ontological commitments of folk psychology. On the one hand, we have the Lewis/Jackson programme or Canberra plan which presents quite clear proposals about how to inquire into metaphysics. It also allows us to be

revisionists about a discourse if it isn't functionally realizable by physical reality. This approach seems to be unsuitable for dealing with the modified communicative conception. On the other hand, we have the pragmatist approach which is clearly more in line with the pretensions of the communicative conception. But this position isn't able to present any clear guidelines for doing ontology besides remaining entirely quietist about that topic. Of course, quietism is also an answer to ontological questions but it seems like a too superficial solution to them.

Since I've adopted the pragmatic approach already, and my defense of the modified communicative conception has assumed it as an overall framework, what I am now going to do is to argue that the Lewis/Jackson programme is a problematic approach and that ultimately the pragmatic account (in its somewhat modified form) is preferable. After that we can inquire into whether the latter needs to remain committed to agnosticism and quietism about ontological matters, something that I take as a default assumption for now.

## **6.4 Against the Lewis/Jackson view**

As we saw there are two main steps in the Canberra plan. First, specify the folk theory that a concept implies. Second, see what kind of functional roles need to be fulfilled in order for this theory to be true. The additional step, which includes seeing whether these functional roles can be realized by structures that a physicalist can find acceptable, doesn't belong, strictly speaking, to specifying ontological commitments of commonsense. It rather belongs to the clarification of whether these commitments can be vindicated and I am not tackling this step at this point. Also, I won't deal with the second stage of that programme which is the technical issue of how to interpret theoretical statements. My criticism will focus entirely on the first step which, if not feasible, undercuts the whole programme.

The path I am *not* going to take is that of casting doubt upon the notion of conceptual truth, so that no inference is seen as essential to concept-possession (à la Williamson 2006). It is true that Jackson's approach, for instance, assumes the feasibility of that notion. But I am not going to question this idea directly. Rather, I will argue that the assumption that intuitive judgements of the folk reveal an implicit theory is already problematic, whether or not that theory is supposed to consist of conceptual truths. The idea is that there is no clear-cut criterion for deciding which assumptions form the putative theory. For this reason, the first step of the Canberra plan can't get off the ground.

So why can't the intuitive judgements of the folk concerning commonsense psychology reveal the folk theory? One reason for this is the fact that it is unclear what kind of judgements would settle this issue. Should we aim at the unreflective judgements of the man on the street or should we prefer the reflective judgement of a philosopher? It isn't clear which judgement reveals the semantic competence of individuals. Choosing either of these possibilities

will spell trouble for the investigator. If we prefer unreflective judgements, we disparage the role of reflection in our conceptual competence. If we focus on reflective judgements, we are threatened with the possibility that reflection distorts the process of evaluating the application of the concept. Which kind of judgement reveals the putative theory that characterizes the concept? This question can't be answered without bringing in strong assumptions about what should be counted as the right kind and it isn't clear how this can be done. To make matters worse, there is probably a continuum between unreflective and reflective judgements. One can distinguish between more or less immediate and more or less rationally mediated responses. One could hope that people end up with the same judgements both in reflective and unreflective cases. This is a highly dubious assumption. As long as we consider the judgements of philosophers, opinions about what assumptions the concepts of belief and desire involve are highly divergent.

I actually think that reflective judgements should be given preference when one does philosophy.<sup>106</sup> But I say this from the perspective of the pragmatic approach – an approach which isn't interested in probing people's intuitions in the first place. Because of that it doesn't face the same problem. At least I hope to show that it doesn't. The aforementioned dilemma arises for the Canberra plan, though, because the latter is aimed at revealing the folk theory that is supposed to belong to the semantic competence of persons.

A second consideration against the Lewis/Jackson programme is that it is possible that the ability to make judgements about concept-application isn't enabled by a theory at all. It is an open question whether theory theory is an approach that we should adopt in order to explain this ability. There are numerous alternative accounts on the market at the moment. Some of them deny that there is any role for a theory at all (Gallagher 2006; Hutto 2008). Most of them take theory to be only one element among many that contributes to the explanation of folk psychological understanding (Goldman 2006; Sitch & Nichols 2003; Currie & Ravenscroft 2002). I won't delve deeper into this debate. It is important to note that even if a theory plays some part in explaining the abilities of the folk there are other mechanisms besides that. This already gives us a reason to think that whatever theory we can eventually glean from those studies reflects the folk conceptual competence only incompletely. It thus gives us a distorted or limited view of the concepts that we are interested in.

My third argument against the feasibility of the first step of the Canberra plan is that even if we are able to identify which folk judgements are appropriate in this, it is unclear whether they should be given the authority to decide our ontological commitments. The first two arguments assumed that if the folk theory is explicable we have a trustworthy starting point from which to

---

<sup>106</sup> “How can it be a definitional truth that spontaneous judgements as opposed to careful and reflective judgements are particularly effective for those determining whether an object is in the extension of a concept?” (Cappelen 2010: 210) Cappelen raises this point when he argues against Goldman.

inquire into what the folk are committed to. But even if we can discover such a theory, should it be taken as a satisfactory reflection of what the folk are actually committed to when they ascribe mental states to one another? After all, conceptually competent people might be mistaken in their judgements about concept-application if these judgements require thinking about complex counterfactual situations. It is possible, for instance, that even if there is a common pattern in the majority of folk judgements, this doesn't really reflect folk commitments. It might be an outcome of a particular interpretation of what the folk say instead. This is something that Hutto brings out with the example of how the supposedly intuitive idea that mental states need to be causally efficacious in a strong sense may be the result of a contemporary scientific picture (Hutto 2011). According to this picture everything obeys causal laws, paradigmatically understood along the lines of the natural sciences.

So it is possible to question, despite the folk intuitions, whether we are committed to the idea that beliefs and desires are inner causes of behaviour. This idea might be an outcome of the fact that mental states are often modeled on physical causes (see Ryle 1949, Ch. 1). Let's just consider another example. It may be intuitive to judge that memory functions as a storage capacity. Remembering that  $p$  would then involve storing a memory trace, a representation that  $p$ , in the brain since the time when this representation that  $p$  was originally acquired. Despite that, if Wittgenstein is right and remembering something doesn't require instantiating continuous memory traces, should we say that the folk conception of memory has such a commitment (Wittgenstein 1980: 220; see also Moyal-Sharrock 2009)? One could, of course, dismiss an alternative conception of memory because it contradicts the folk view and doesn't reveal actual folk commitments. But why do we have to prefer such a judgement to others? We should at least leave open the possibility that the Wittgensteinian view captures those commitments. The storage view could be an outcome of a confusion that arises from the use of metaphors in descriptions of memory. After all, Wittgenstein also belonged among "the folk", having been an adult member of *homo sapiens*. The storage-metaphor may be very common even cross-culturally (Amberber 2007). However, this doesn't mean that we have to rely on this metaphor in order to sort out our ontological commitments.

The point of these examples is that folk judgements about the conditions of concept-application need not always be trusted, at least in situations when we can interpret their use of concepts in less loaded terms. The question here is: under what conditions does a claim about a conceptual commitment have authority? When should we accept a putative commitment? It is doubtful whether authority can be conferred on a claim only because the majority of people tend to endorse it spontaneously. Let's take a more radical example. One of the most common assumptions about beliefs and desires that is also taken as a crucial element in the folk theory is the law that is modeled on the practical syllogism. This law doesn't have a canonical formulation but one of the most usual ways to put it is as follows: if a person wants that  $p$ , and believes that

doing X will bring about  $p$ , he will, all other things being equal, do X (cf. Gopnik & Meltzoff 1997: 126). Or take a more elaborate version of this law: if a person wants that  $p$  and believes that *if  $q$  then  $p$* , and if she is able to make the case that  $q$  then, excluding conflicting desires, she will bring about that  $q$  (Churchland 1981: 71). A proponent of the Lewis/Jackson account could claim that the folk are committed to the belief-desire law in their mental state attributions. The law would then specify a necessary condition for someone to have beliefs and desires.

But even this widely recognized case of a putative implicit assumption that governs our folk psychological attributions doesn't have to be a guide to the ontology or constitutive properties of mental states. Although this assumption doesn't seem to be an outcome of an implicit metaphor or a metaphysical picture, it isn't something that we must treat as an inevitable component of the folk concept. At least this further move requires an argument. Why should we think that if many people find this law to be intuitive they must be committed to it? To find out such a commitment we should do more than probe their intuitions. We should observe how this concept is actually applied and what it provides us with. The outcome of this may well be that we find ourselves asking about intuitions concerning the application-conditions of the concept in hypothetical situations that are completely removed from any everyday context. In these conditions the answer might become too indeterminate (for an elaboration of this danger, see Baz 2012).<sup>107</sup>

So there are serious reasons to doubt that the Lewis/Jackson approach to ontological commitments will deliver the goods. There are problems with specifying the right kind of intuitive judgements in order to fix the essential assumptions of folk psychology. It is also problematic to assume that folk psychological judgements are subserved by a theory. Finally, one should not take intuitive judgements to have final authority over the ontological commitments of folk psychology. This is good news for the modified communicative conception because the latter didn't fit well into the Canberra plan anyway. Unfortunately the alternative view also has its problems.

## 6.5 Problems with the pragmatic view and the solution

Although there are serious problems with the Canberra style of philosophizing, the pragmatic approach I have envisaged is not without difficulties either. Aside from not providing any clear instructions for identifying the ontological

---

<sup>107</sup> Considering that I expressed doubt about the predictive potential of folk psychological concepts (Chapter 2), one might raise suspicion that the reason why I don't want to commit the folk to belief-desires is that I consider the latter to be mistaken. The folk can be committed to something false, I don't deny that. Here the aim is simply to argue for being skeptical about the prospects of sorting out what those commitments really are by relying on people's intuitive judgements.

commitments of the mentalistic vocabulary, there are also more imminent methodological issues that I need to address.

We can clearly do various things by attributing mental states – negotiate and manipulate the attitudes of the attributees, simply phantasize about others, explain and predict their behaviour<sup>108</sup>, etc. The natural question arises how can we constrain this diversity so that we can see what actions are essential to folk psychological attributions and which can be dismissed as being peripheral to the commonsense practice. For instance, Hutto's pick from the domain of different doings that folk psychology involves focuses on those occasions when we make rational sense of other people's actions in terms of beliefs and desires. These explanations are not formulated in the form of general laws but they are responses to why-questions concerning particular behaviours (Hutto 2004; 2008, Ch. 1). If that is true, the folk aren't committed to there being general psychophysical laws. On the assumption that causality implies such general laws, one could draw the further conclusion that the folk are not committed to the causal efficacy of beliefs and desires.<sup>109</sup> That would be an ontologically significant conclusion, based on the pragmatic approach.

We might ask, why is it necessary to pick out a specific element of folk psychological attributions and disregard the rest? As already noted, there are other things we do with folk psychological attributions. The additional difficulty is that actions can be described in multiple ways. This is demonstrated by Anscombe's well-known example of a person pumping water into a cistern but also poisoning the water supply and killing the inhabitants in the process (Anscombe 2000: §23). So even if we admit with Hutto that normalizing explanations are the most salient actions that folk psychological attributions involve, there still remains the problem of justifying the dismissal of other possible descriptions that apply to those doings.

The problem of specifying the central elements of folk psychological attributions doesn't uniquely characterize the pragmatic view. Remember that the Jackson/Lewis approach faced similar difficulties with determining the central postulates of folk psychology. Nevertheless, it at least gave us a possible means for explicating the folk theory (viz. by probing intuitions). I argued that this method isn't feasible but the question of how to identify the central postulates is left open entirely in the case of the pragmatic approach. Unless we want to fall back again on intuitions, there should be another way to specify the criteria for picking out the ontological commitments of folk psychological attributions.

I suggest that the modified communicative conception that I have defended can achieve this. If we want to decide which are the ontological commitments of folk psychology then we should ask whether folk psychological practice can

---

<sup>108</sup> Note that by arguing against the centrality of the predictive function of folk psychology in Chapter 2 I have not said that people never try to predict behaviour by ascribing beliefs and desires.

<sup>109</sup> In fact, Hutto (1999) has himself argued for the view that reasons are not causes.



still work if we don't assume the putative commitment in question. So we must ask whether our practice of attributing beliefs and desires still makes sense after we have given up that putative commitment. Only those assumptions of mental state attributers that really are required for the practice to make sense need be considered as essential commitments of folk psychology. Note that this is quite different from OWE. The point here isn't that beliefs and desires have a deeper essence for natural sciences to discover and that thanks to this we can dismiss many of the folk assumptions. The idea here is that the essential aspects of folk psychological attributions are revealed when we consider which requirements need to be met if attitude ascriptions are to succeed.

The explication of the concepts of belief and desire in the previous chapters also helps us to understand the ontological commitments of mental state attributions. The explication considered the mental state attributions in abstraction from the messy implications they can have in our everyday life. What were important were the main benefits that the folk psychological attributions offered. The explication brought out the motivation for introducing the concepts of belief and desire in the first place in a way that would explain those benefits. These concepts were meant to extend the practice of communication and explain certain kinds of communicative behaviour, thus articulating the minimal conditions under which the mental state attributions make sense for the attributers. The lesson from the explication was that the attribution of desires minimally requires the ability to command on each others' behalf and a sensitivity as to whether those vicarious commands would provide subjective satisfaction for those on whose behalf the commands are uttered. The attribution of beliefs, on the other hand, requires the ability to utter vicarious assertions and a sense of which behavioural dispositions are relevant for doing that. If these abilities are the minimal requirement for possessing the concepts of belief and desire then we can also define the most crucial assumptions of mental state attributions: that the possessors of beliefs and desires are capable of having behavioural and hedonic dispositions and of participating in the practice of linguistic communication. These assumptions reveal what attributers have to commit to if the attributions are to deliver their benefits, even if the folk intuitions involve more. One can disregard the putative commitments to the causal efficacy of attitudes and specific internal structures in agents' brains, for instance, because the attribution of attitudes delivers its benefits also without those commitments. The explication thus realizes the pragmatic approach by extracting the commitments of folk psychological attributions from the (basic) use of the concepts of belief and desire.

The present proposal is somewhat different from the other pragmatic approaches. It doesn't merely ground the commitments of folk psychology in everyday concept-use. It also has a specific method of specifying which uses are relevant for articulating folk commitments. This method relies on the benefits that folk psychology offers. Hutto also picked out a particular use of mental state concepts which he considered to be relevant (Hutto 2008). But he seemed to think that in order to justify his choice it would suffice to note that this is

what we do. I have not settled with that. The commitments of the modified communicative conception can be established even if we can't identify the main use of the concepts of belief and desire by merely observing folk psychological practice. We have to consider the benefits of this practice too.

From the minimal assumptions about attribution-conditions one can ask what beliefs and desires have to be like in order for these assumptions to be true. In the next section I will address the possible responses to that question and argue for my own realist and dispositional account. Before going on, I should mention that by relating the ontology of beliefs and desires to the minimal assumptions one has to make about them, MCC isn't a form of agnosticism about the metaphysical commitments anymore. The aforementioned minimal assumptions are such that one is committed to them if one is to make sense of the concepts of belief and desire. The question of quietism, on the other hand, will be answered at the end of the chapter.

## **6.6 The ontological status of beliefs and desires**

### **6.6.1 Initial proposal: interpretivism**

To find an answer to the question "what does having beliefs and desires consists in?" (presuming that the modified communicative conception is correct) perhaps the most promising place to look is the original communicative conception as defended by Gauker. In his 2003 book he distinguishes between three different ways of explaining the nature of a thing: 1) by characterizing its internal structure; 2) by identifying its function; 3) by finding its place in human conventions (Gauker 2003a: 272). In the case of propositional attitudes the first would be the approach of those who support Lycan's intuition and think that there is an essence of beliefs and desires for science to discover. The second would be the approach of the Canberra planners who take the functional profiles of belief and desire to characterize their natures. Finally, the third view is the one Gauker in fact adopts. If we view beliefs and desires in such a way, their nature is analogous to the nature of weddings, funerals, and all those things whose nature is exhausted by their being socially recognized as having such a nature. The communicative conception in its original form, then, is an interpretationist position, and Gauker himself adopts the interpretationist label (Gauker 1994).

Before evaluating whether this move is successful we should ask, what does interpretivism as a philosophical position actually involve? It is most closely associated with certain philosophers, most notably Dennett and Davidson. The starting-point of Dennett's interpretivism is his distinction between three stances by means of which we can explain and predict worldly phenomena: the physical, the design, and the intentional stances. The first involves the use of physical laws. The second consists in the attribution of functions to the system on the assumption that it functions properly. The intentional stance involves the attribution of beliefs and desires to a system on the assumption that that the

latter is rational and predicting its behaviour on the basis of that. If the behaviour of a system can be interpreted by taking an intentional stance towards it, the system has beliefs and desires (Dennett 1987). So interpretivism in the case of Dennett means that being interpretable from the intentional stance is sufficient for something to have beliefs and desires. This allows also thermostats and chess-playing computers to have those attitudes if interpreting them by using the intentional stance is predictively useful.<sup>110</sup>

Davidson's interpretivism is tied to his theory of radical interpretation which postulates a hypothetical situation of learning a language from scratch, relying only on the behaviour of the language-users. This process also involves the attribution of beliefs and desires to them. Having those attitudes consists in being interpretable in that way (Davidson 1973). Notice that Davidson is pushing towards the idea that interpretability is a necessary condition for having propositional attitudes when he claims that "what we could not interpret is not thought" (Davidson 1990: 14).

The main idea of interpretivism, then, is that having beliefs and desires stands in some constitutive relation to our practice of attributing beliefs and desires. As Alex Byrne puts it: "There is no gap between our best judgements of a subject's beliefs and desires and the truth about the subject's beliefs and desires" (Byrne 1998). I take this to be a minimal assumption that should characterize any version of interpretivism. Gauker's idea that beliefs and desires are social kinds fits nicely into this characterization. After all, it is natural to be an interpretationist about social kinds. Otherwise we would have to make sense of the idea that weddings and funerals could exist without anyone treating them as weddings and funerals.

We can go even more minimal in our understanding of interpretivism than this. For William Child, for instance, interpretivism could involve just the claim that we can come to understand mental states if we consider the interpretation by which they are ascribed (Child 1994: 9). This kind of conception doesn't entail that interpretability is necessary or sufficient for having mental states. It is more like a methodological claim than an ontological view. Yet Child is a bit vague on this matter because elsewhere he claims, like Byrne, that a fully informed interpreter can know everything there is to know about someone's thoughts (Ibid. 12). This would amount to the claim that interpretability is sufficient for having thoughts.

Interpretivism is not then uniformly understood and there is a continuum between stronger and weaker interpretations. The strongest position claims that interpretability is both sufficient and necessary for having beliefs and desires. Positions that admit either only a sufficiency or a necessity requirement are in the middle. The weakest form of interpretivism makes only a methodological claim: looking at the facts about interpretation is supposed to provide insight about the objects of interpretation. Let's see whether the communicative

---

<sup>110</sup> It is unclear whether Dennett's theory is justified to claim that the interpretability is a necessary condition, too.

conception could fit onto this continuum and whether a marriage of the communicative conception with interpretivism makes sense. Since methodological interpretivism is the most minimal position, I take it as the last opportunity for interpretivism to retain any substance. For this reason, I consider first the more substantial versions of that position.

So the first way of conceiving the communicative conception as a form of interpretivism is Gauker's suggestion that beliefs and desires can be taken as social kinds. This would be a quite strong form of interpretivism because it can be read as a claim that the social recognition of something as having beliefs and desires is both necessary and sufficient for it to have those attitudes. This view faces the immediate threat of falling back into a more realist position because it still needs to say what the nature of social kinds is. If the latter can be reduced to psychological states, perhaps even to beliefs and desires, then the communicative conception wouldn't be an interpretationist position. Even worse, it would be circular because it presupposes what it explains.

Even if this worry can be fended off, I am still very skeptical about this way of understanding the nature of attitudes. Nothing in their definition gives one a reason to equate beliefs and desires with social kinds. It is true that their attribution is understood in the context of social interaction. But if attitudes are social kinds, their instantiation should be dependent on the existence of the practice of recognizing them as beliefs and desires. It is difficult to accept such a strong form of ontological dependence because it makes sense to say that individuals can have beliefs and desires even if a practice of interpretation doesn't exist. It is highly plausible to think that there would be no weddings if no practice of taking something as a wedding existed. The same can't be said about beliefs and desires. One can conclude, then, that social kind interpretivism gets wrong the necessary conditions for having beliefs and desires. This concern is even more pressing for the modified communicative conception which made more assumptions about the ascription-conditions of beliefs and desires than the original communicative conception. These assumptions concerned hedonic and behavioural dispositions. Since the attribution of attitudes is meant to respect such facts about the attributees, it is even harder to see how those attitudes could be identified with mere social kinds.

Locating beliefs and desires among the social kind categories isn't really the standard way to formulate interpretivism. An alternative definition that relies heavily on the idea of an ideal interpreter can be found in Byrne's (1998): "x believes that  $p$  iff if there were an appropriately informed Ideal Interpreter, she would be disposed to attribute to x the belief that  $p$ ." This kind of treatment of interpretivism (let's call it Ideal Interpreter Interpretivism, or III) could be applied to desires as well. Child also seems to assume some kind of ideal interpreter when he claims that what actual interpreters learn about thoughts isn't all there is to learn but what they *could* learn is all there is to learn (Child 1994: 30). The communicative conception could also be interpreted through that lens. On such an interpretation the communicative conception claims that the ideal interpreter can determine what can be asserted and requested on a person's

behalf, thus determining what the person believes and wants. This fares better than social kind interpretivism because the ideal interpreter need not be actual. Even before interpreters were available, individuals could have had beliefs and desires because the ideal interpreter would have attributed these attitudes to them.

The problem with this kind of approach is that it requires explaining what makes the ideal interpreter an ideal one, without relying on the specification of interpretation-independent facts. The most obvious way to sort out what being an ideal interpreter consists in is to specify what is the nature of the *objects* of interpretation what the ideal interpreter is capable of understanding. The success of the ideal interpreter could then be explained by her ability to consider the facts about mental states. But that would mean that we can explain the nature of the objects of interpretation independently of facts about interpretation. The interpretation couldn't determine what it is to have mental states.<sup>111</sup> One move that a defender of III could make is to claim that although the previous argument shows that the facts about mental states are explanatorily prior to facts about interpretation, the second are still ontologically prior to the first. But that would be a rather surprising move because no reason is given as to why we should accept it. It seems like a desperate attempt to save interpretivism, not a well-argued way to articulate the metaphysical dependency-relation between interpretation and mental states.

Another, more promising way to respond to this concern would be to argue that interpretivism can be retained when we confine the facts that are relevant to the ideal interpreter to those that are behavioural. If the explanation that makes an ideal interpreter ideal doesn't make any reference to mental states then III can still be an informative metaphysical position. It would explain the facts about mental states in terms of facts about the ideal interpreter and non-mental facts. It is noticeable that interpretivism in general seems to be motivated by the goal of understanding mental states without taking them to form an irreducible ontological category. This doesn't mean that non-mental facts can't be used to account for mental facts. Taking this into account, ideal interpreter interpretivism can at least avoid the charge that explaining the ideal interpreter requires relying on facts about mental states.

I still have a nagging concern at this point. Although this move to defend III is feasible, it eventually blurs the distinction between interpretivism and behaviourism. It is noticeable that one important motivation to defend interpretivism comes from epistemology. It is motivated by the need to account for the fact that people can have knowledge about attitudes without knowing anything about the internal structure of the brains of individuals. "Interpretationism aims to make it intelligible that an ordinary interpreter can, in favourable circumstances, tell what someone believes and desires." (Child 1994: 47) This means

---

<sup>111</sup> In a somewhat analogous manner, Kripke rejects a defense of the dispositional account of rule-following which relies on idealized conditions by arguing that specifying the ideal conditions already assumes what rule is being followed (Kripke 1982: 28).

that the distinctiveness of interpretivism as a philosophical position eventually comes from what it opposes. For instance, it does assume that sub-personal information processes do not play or do not need to play any role in fixing the facts about mental states. It is the interpretation of the observable behaviour that does this (Child 1994: 31). But if that is true, I don't see why we need to take III to be something different from behaviourism. After all, it needs to take into account behavioural facts in order to explain what makes mental state ascriptions true. The ideal interpreter doesn't seem to play any substantial role in accounting for the truth of interpretations. Her capacities should still be explainable by the facts about the objects of interpretation. The kind of interpretivism that relies on the idea of the ideal interpreter, then, doesn't seem like a very interesting position. But we can still ask whether the communicative conception fits into this category.

Understanding the ideal interpreter only in terms of the facts about interpretation, not about the objects of interpretation, may perhaps be possible when we adopt the original communicative conception. It is possible only because the original communicative conception says very little about the conditions under which something is assertable or requestable on someone's behalf. The modified communicative conception, on the other hand, treated assertability and requestability conditions as a serious problem and specified them in terms of interpreter-independent facts – hedonic and behavioural dispositions. One could object that I *did* allow that mere assertions and requests on someone's behalf are sometimes acceptable. But I've also stated that the practice of belief and desire attribution would not make sense if the dispositions and hedonic states were never taken into account. Because of that, the interpreter-independence is an unavoidable element in the specification of when it is appropriate to interpret someone in terms of beliefs and desires.<sup>112</sup>

The third and the most recent way to articulate interpretivism is Bruno Mölder's ascription theory. This position tries to steer away from the notion of the ideal interpreter and relies on the idea of canonical interpretation. This is understood in terms of our everyday conception of mental states which is available to actual interpreters. Nevertheless, the ascribability of mental states can often transcend the actual attributions which don't always follow the conditions of ascribability. In this way, it allows for errors in interpretation (Mölder 2010: 172–173). Ascription theory, then, seems to exist somewhere in the middle between social kind interpretivism and ideal interpreter interpretivism. It assumes that the folk have the competence to discern which are the correct ascribability conditions but they may not always perform in accordance with this competence.

Another aspect of ascription theory that distinguishes it from other versions of interpretivism is its reliance on the idea of pleonastic properties in its

---

<sup>112</sup> III faces an additional problem. It seems plausible that we could adopt III towards almost any area of discourse. If the ideal interpreter is really the *ideal* one, she should basically be all-knowing which means, of course, that she would know all the facts.

articulation of the metaphysical difference between the mental and the physical. According to Mölder, mental properties such as having a belief that  $p$  and having a desire that  $p$  are pleonastic properties. This means that whether something has them is wholly dependent on it falling under a corresponding concept of that property (Mölder 2010: 23). The idea of pleonastic properties, taken from the work of Stephen Schiffer, is about those properties that fall under pleonastic concepts. A pleonastic concept is a concept the application of which entails that the corresponding property is instantiated. Such a property is only instantiated in a deflationary sense which means that it isn't a natural property. It doesn't belong to the "furniture of the world" (Schiffer 2003: 57). The ontological status of mental states is fully exhausted by the application of the corresponding predicates. Since ascription theory takes mental states to be pleonastic it doesn't allow for a more substantial story about those states than the story about the application-conditions of the corresponding predicates.

Mölder's ascription theory is certainly more articulated than the previous two. It is also a rather strong version of interpretivism. I take it that it pretends to give both necessary and sufficient conditions for having mental states. Both can be specified in terms of canonical conditions of interpretation. "Which mental properties one has depends solely on which mental predicates are ascribable to one." (Mölder 2010: 40) A quite obvious strength of ascription theory is that it eschews the notion of the ideal interpreter which, as we've seen, doesn't do any interesting metaphysical work. It still faces the same question as the ideal interpreter interpretivism did: how to specify ascribability conditions without making interpretation dependent on the facts about the objects of interpretation or falling into behaviourism. Of course, interpretationists need not deny that there are non-mental facts that the interpretation can depend upon. As we've seen, though, this dependence puts the metaphysical weight on the facts about behaviour and behavioural dispositions, not the interpretation itself.<sup>113</sup>

The ascriptivist can call the notion of pleonasticity to rescue. She could try to salvage the distinctiveness of interpretivism by noting that, unlike behaviourism, beliefs are not inflationary properties that can be identified with behavioural dispositions. They are deflationary entities and their ascription only *takes into account* such dispositions. Also, ascription theory can allow that other physical properties function as the ascription-base, without implying that beliefs and desires reduce to these properties. It seems that even from the perspective of the modified communicative conception, beliefs and desires can still be seen as pleonastic. Their ascription relies on certain psychological (hedonic) and behavioural facts but not any facts about the attitudes themselves. But now it seems that interpretation has lost its constitutive role in the characterization of ascription theory and that the metaphysical work is done by the pleonasticity. Why not simply call this position anti-realism according to which there are no

---

<sup>113</sup> The dependence of the interpretation on the facts about behavioural dispositions doesn't mean that these dispositions have to be fully characterizable. It simply means that there are such dispositions.

genuine facts about beliefs and desires and these attitudes are merely pleonastic entities?

Of course, if pleonasticity is defined in terms of interpretability then Mölder could argue that an element of interpretation does play an irreducible role in his theory. I grant that, but the objection still stands that the fulfilment of canonical interpretation-conditions requires explanation. This explanation needs to make reference to facts about the objects of the interpretation. If the latter are merely pleonastic and interpretation-dependent, then it is hard to make sense of their explanatory power. Ascription theory may retain its distinctiveness thanks to the appeal to pleonasticity but it still faces difficulties when explaining how and when interpretation-conditions are fulfilled.

All three versions of interpretivism, then, have some noticeable problems: social kind interpretivism blurs the distinction between mental states and social kinds; ideal interpreter interpretivism has troubles explaining what makes the ideal interpreter an ideal one; and the ascription theory seems to boil down to mere anti-realism about the mental. But we can still take the original formulations of interpretivism as a starting-point and ask whether the modified communicative conception is committed to the claim that interpretability is sufficient and/or necessary for the attribution of beliefs and desires in the first place. Furthermore, we can ask whether the modified communicative conception is committed to anti-realism about the mental.

### **6.6.2 Interpretability and the modified communicative conception**

The problems with defending interpretivism appeared to be quite pressing. But let's have a look at whether the modified communicative conception even needs to be committed to some version of interpretivism as Gauker seems to presume. It does seem that the proponent of MCC should accept methodological interpretivism. I've maintained that investigating how we ascribe beliefs and desires to each other is the way to understand the nature of these attitudes. As we've seen, William Child was actually quite sympathetic to that kind of interpretivism. Yet it doesn't seem to be a very informative position. After all, a behaviourist or a functionalist could also be a methodological interpretationist in this sense. Another weak version of interpretivism would be the claim merely that beliefs and desires do not entail anything specific about the internal properties of agents (Mölder 2010: 40). In that case the modified communicative conception *is* a version of interpretivism. But understood in such a way, it doesn't say anything distinctive about the ontology of beliefs and desires because a behaviourist could state the same.

As for stronger versions of interpretivism, the modified communicative conception doesn't entail that assertions and requests on a person's behalf are sufficient conditions for that person to have beliefs or desires. The facts about hedonic and behavioural dispositions have to be in place, too, and these facts determine ascribability-conditions, not the other way around. The dependence of correct interpretations, even ideal interpretations, on facts about the objects



of interpretation is an assumption that is very difficult to avoid. The sufficiency-claim is thus a non-starter for understanding MCC as a form of interpretivism.

The necessity-condition is more plausible because without that the communicative conception would seem to be indeterminate between different theories on the ontological level. It seems, then, that the modified communicative conception could retain its interpretationist credentials, at least when we take interpretability to be a necessary condition for having beliefs and desires. This interpretability is not understood in terms of an ideal interpreter but in terms of conditions that are set by our shared folk psychological ability to assert and request on the behalf of others.

But what do I mean by the claim that without the necessity-condition the modified communicative conception would be indeterminate on the ontological level? The answer is that if interpretability weren't the necessary condition for having beliefs and desires, MCC would seem to allow for almost any kind of metaphysical theory about the mental. But the distinctiveness of MCC can be retained without taking it to be *ontologically* distinct. It is, in the first instance, an account of the ascription of beliefs and desires, not an account of their nature. Its distinctiveness is wedded to the former, not the latter. Although it makes only minimal ontological commitments, it doesn't entail that these minimal assumptions exhaust the nature of attitudes. So if the only motivation for adopting interpretability as a necessary condition is that otherwise MCC would be an ontologically indeterminate position, this motivation doesn't amount to an argument for the interpretability being a necessary condition. And I am skeptical about whether the modified communicative conception itself can provide that argument. So we seem to be stuck where we began: the communicative conception doesn't entail anything specific about the nature of propositional attitudes.

But maybe it's a good thing. After all, we have seen from the discussion about different interpretationist theories that they have serious problems. If the modified communicative conception isn't committed to interpretivism then it doesn't face the same challenges. That still leaves open the matter of which metaphysical view of propositional attitudes we should adopt. In what follows, instead of going through different theories and analysing their strengths and weaknesses, I will propose what I consider to be the most promising account. Perhaps surprisingly it relies largely on Daniel Dennett's theory without being interpretationist in any interesting sense.

### **6.6.3 The communicative conception and motivations for interpretivism**

For the purposes of laying out my own proposal, it will be useful to take interpretivism as a starting-point. It reveals two motivations which help us see in which ways theorizing about the mental can go wrong. I suspect that the main motivation for interpretivism is that mental properties are taken to be

metaphysically more “shallow” than the physical ones – facts about their nature do not in any way transcend the ability of competent concept-users to ascribe them. This also clarifies the reason there is for supposing that mental properties are merely pleonastic and that their existence is dependent on the corresponding concepts being introduced into the discourse. I suspect, then, that the first motivation for interpretivism comes from the implicit assumption that the talk about beliefs and desires isn’t ontologically fully respectable and that we need to give an account of mental discourse that explains how we can still employ it without hurting our rational conscience.

It is doubtful, though, that the modified communicative conception needs to go along with the idea that mental properties are somehow ontologically less respectable than physical properties. It is a superfluous assumption. When we consider how it is determined whether someone has certain beliefs or desires (whether certain assertions or requests can be uttered on her behalf), there seem to be facts of the matter about those mental states. Mental properties are not really ontologically different from such properties as height or weight, for instance. One main idea behind the notion of pleonastic properties is that these are properties which are introduced by “something-from-nothing transformations” through the application of corresponding predicates. This idea is sensible when we consider fictional entities which exist in only a deflationary sense. But to say the same about mental properties seems to be quite ill-motivated. After all, those who assert and request on the behalf of others need to take into account coarse-grained facts about others’ dispositions in order to count as successful belief and desire attributions.

What might give the impression that the modified communicative conception is committed to the view that mental properties are pleonastic is that it doesn’t identify beliefs and desires with any particular physical property. It is true that beliefs and desires are specified in terms of their ascription-conditions. They are not identified with any particular physical or mental property or any complex of them. This might give an impression that by avoiding the question of what beliefs and desires are I have committed myself to a form of anti-realism about propositional attitudes. This is the idea that, strictly speaking, there are no interpretation-independent facts about what one believes and wants.<sup>114</sup> But I simply reject that view and maintain that if the application-conditions for mental predicates are established there is a fact of the matter as to whether a person has a given belief or desire.<sup>115</sup>

However, the motivation for anti-realism might not only come from an ontological worry. The second motivation for interpretivism is that it gives an

---

<sup>114</sup> One can also express anti-realism about beliefs and desires in terms of them only existing in some deflationary sense. I find this kind of talk having the uncomfortable implication of making beliefs and desires thing-like (see Steward 1997: 253). Since I don’t think that those attitudes are things that need to be located in some place or time, I prefer to talk about facts.

<sup>115</sup> If this is enough to commit myself to interpretivism, then I am an interpretivist. But considering the distinctions made earlier, I think that this only counts as a version of methodological interpretivism.

answer to the question, “how can we know about others’ mental states?” (Child 1994: 47). Canonical ascribability conditions are available to folk psychologists. Interpretivism offers an easy explanation for our knowledge of mental states because it equates having an attitude with being interpretable by us. It thus contrasts with views according to which to have a belief or desire is to have certain internal states, for instance. Unlike interpretivism, such a view faces trouble when one wants to explain how the ordinary folk who have no working knowledge about neuropsychology are able to know about each others’ beliefs and desires.

My answer is that one can accept the modified communicative conception and also agree that our everyday ascriptive practice allows us to know what people believe and desire. The reason why knowledge of mental states comes (relatively) easy is because of the epistemic accessibility of the criteria of attribution. The epistemology of folk psychology is thus very reliable. But this doesn’t mean that we should be (non-methodological) interpretationists about it. Without the ontological worry there is no need to claim that the facts about mental states are somehow more dependent on interpretation than other facts.

The modified communicative conception is consistent with a realism about beliefs and desires, their knowability to ordinary interpreters and rejecting interpretivism.<sup>116</sup> To see how unproblematic these assumptions are let’s compare the situation here with the challenge faced by a realist in meta-ethics. She is committed to the idea that normative claims are true because of the normative facts. She faces the challenge of explaining how these facts fit into our general metaphysics, how they are epistemically accessible and how they they can guide our deliberations (McPherson 2011: 224). The realist about normative discourse faces problems on every front<sup>117</sup>. The realist about the mental, on the other hand, who also defends the modified communicative conception, can answer them quite easily. Facts about beliefs and desires are facts about dispositional and hedonic profiles of persons and these facts are epistemically accessible to competent interpreters. What’s more, the communicative conception need not be troubled by the third challenge which is only specific to meta-normative theories. MCC, then, doesn’t need to buy into a form of anti-realism in order to hold onto its commitments.

#### **6.6.4 “Realistic Dennett” and the communicative conception**

To make it more explicit how realism about attitudes and MCC fit together it is useful to draw some comparisons with other views. It may be somewhat surprising but the metaphysical position about the mental that I have hinted at

---

<sup>116</sup> It doesn’t reject methodological interpretivism, but as we’ve seen, this is not a very substantial position.

<sup>117</sup> I am not implying, of course, that a realist is inevitably unable to answer those problems. This example is only meant to illustrate the relative ease that a realist about mentalistic discourse has.

here is quite close to Dennett's theory if we interpret the latter in the right way. According to Dennett, the question about the reality of beliefs and desires can be compared with the question about the reality of centers of gravity. The existence of such abstract entities only entails *mild realism* about them (Dennett 1991b: 29). Although beliefs and desires of persons can be taken to be real patterns in their behaviour, there are no deep matters of fact about them if we compare them to the entities postulated in physics, for instance.

Whether those patterns are discernible in the behaviour of a system for an attributer is dependent on an attributer's ability to take the intentional stance towards that system. By intentional stance Dennett means a perspective from which one can predict the behaviour of a system by attributing beliefs and desires to it, on the assumption that it is rational (Dennett 1987: 17). This stance is different from the other two stances: the design stance and the physical stance. From the design stance we predict a system's behaviour by assuming it to fulfil a certain function. The prediction of behaviour from the physical stance relies upon physical laws that are supposed to govern a system's behaviour. Dennett takes the physical stance to be the most respectable perspective from which to investigate real-world systems because they inform us about the basic structure of reality. The epistemic credentials of the intentional stance are relegated to a secondary position. Nonetheless, although the intentional stance doesn't reveal to us the ultimate structure of reality, it provides an economic way to predict the behaviour of complex systems such as humans. The patterns that are discernible from the intentional stance, then, are taken by Dennett to be real in a weaker sense than those that can be discovered from the physical stance. That is why he calls his position *mild realism*.

When it comes to the function of folk psychology there are quite obvious differences between the modified communicative conception and Dennett's account. According to the first, fulfilling the attribution conditions of belief and desire guarantees successful attribution but the success need not be measured by the predictive power of such attributions. Whether or not mental state attributions provide any straightforward predictions of behaviour, their communicative and explanatory significance is the reason why they earn their keep. Unlike Dennett, I would also put more stress on patterns in the linguistic behaviour of persons in revealing their attitudes. In these respects Dennett's theory of the intentional stance is inconsistent with the modified communicative conception. But this doesn't mean that we should throw the baby out with the bathwater. The idea of beliefs and desires as real patterns in people's behaviour is something that can still be taken on board. It makes sense to talk about real patterns even if we can't make any specific predictions that rely on them. As long as a person's behavioural and hedonic dispositions conform to patterns that are characteristic to beliefs and desires, that person also has those attitudes. The precise characterization of those patterns may be too multifarious to be fully descriptively captured but we have the ability to discern them. It is rather more useful to rely on paradigmatic cases of such patterns. For instance, a stable disposition to assert and request reveals a pattern that is characteristic to the

corresponding belief and desire but this can be complicated by additional patterns in behaviour that force us to withdraw the attribution.

The adoption of pattern-talk from Dennett doesn't necessitate becoming a mild realist and taking beliefs and desires to have less than a respectable ontological status. As Child points out, Dennett's theory of beliefs and desires actually doesn't need to make any essential reference to interpretability in order to account for those attitudes (Child 1994: 51). If we give up the idea that some patterns are more real than others, the talk of *mild realism* loses its point (Hutto 2013). For instance, the fact that mental state attributions don't say anything specific about patterns in the brain doesn't mean that one can't be a realist about those attitudes. Dennett seems to assume that this fact already gives us a reason to adopt mild realism (Dennett 1991b: 42). However, this is a muddle. If something is internal to the brain, it isn't more real than something that isn't.<sup>118</sup>

One reason I can think of as a genuine motivation for anti-realism comes from concerns about causality. One could argue as follows. Since we need to assume the causal closure of the physical and every physical event has a sufficient physical cause, mental events need to be identifiable with physical events in order to be causally efficacious (for a classic presentation of this argument, see Kim 1998). But if this identification is not possible, an interpretationist solution can be that mental events need not compete with physical events because the properties that they instantiate are metaphysically deflationary. By taking mental properties to be merely pleonastic an interpretationist could retain the causal closure of the physical but still talk about mental states in a deflationary sense.

The concern about causal efficacy of the mental or the lack of it also needs to be addressed by the communicative conception. But notice that the version of it that I have defended (as well as Gauker's original account) doesn't involve the assumption that the attributers of beliefs and desires are committed to any strong assumptions about the causal efficacy of propositional attitudes. This assumption is related to the idea that the attribution of beliefs and desires is first and foremost driven by predictive interests. If we reject this idea, the problem of causal efficacy is less pressing.

Nevertheless, I have retained the idea that the attribution of beliefs and desires provides some explanatory understanding of others. This seems to require that we respond to the question: how do beliefs and desires explain behaviour? However, the how-question helps us out of this difficulty. Remember that attributing a belief is a way in which we can explain how an assertion on another person's behalf is acceptable, considering her dispositional behaviour. The attribution of a desire is supposed to explain how a request on another's behalf is appropriate, taking into account her possible hedonic states.

---

<sup>118</sup> Kenyon (2000) seems to think that the fact that it makes no sense to say that there true ascriptions that are unknowable in principle entails anti-realism. But I don't see how it necessarily follows. It simply means that the ascriptions track patterns that are accessible to the ordinary folk.

Are these even causal explanations? My hunch is that they aren't. Rather, mentalistic explanations function like forms of clarification or ways of making sense of other persons. They help us to understand their actual and possible communicative behaviour in relation to their behaviour and pains/pleasures (see Tanney 2009). So it is possible to argue that the worries about the causal efficacy of the mental are ungrounded because folk psychological attributions don't have any commitments concerning the causal efficacy of beliefs and desires.<sup>119</sup>

One might raise the concern that without providing a story about how mental state concepts refer to entities that can be precisely individuated (preferably with reference to certain physical states) talk of beliefs and desires capturing certain patterns in behaviour remains unacceptably vague. I do not think that this is true, though. Why doesn't the story about application-conditions suffice as an account of the nature of beliefs and desires, as long as it is the everyday conception that we are interested in? The feeling that there has to be a more substantial story might stem from the idea that beliefs and desires are particular entities lodged in the mind-brain. Yet it is doubtful that this characterization even makes sense. Helen Steward has relied on grammatical considerations to extensively argue against the idea that there are such things as token states of belief and desire. For instance, explanations of behaviour that cite the attitudes don't seem to function as singular causal explanations that refer to particular events. Rather, they are sentential explanations that appeal to relations between facts and it is the facts of believing and desiring that ground the truth of mental state attributions (Steward 1997: 253).<sup>120</sup> Whether a mental fact holds in a particular case depends on the applicability-conditions of mental predicates being met, not on reference to some wordly particular in the mind-brain. And, in keeping with my account, that doesn't require that the fact of believing and desiring is interpreter-dependent in any interesting sense.

### 6.6.5 Comparison: Schwitzgebel

An illustration of the kind of ontological attitude I am advocating is Eric Schwitzgebel's dispositional account of belief. Schwitzgebel identifies beliefs with clusters of dispositional properties which are associated with belief's dispositional stereotype. Unlike in more standard dispositional accounts, this stereotype also includes dispositions to feel and experience in certain ways (Schwitzgebel 2002: 250). In his later work he associates his conception of belief with what he calls a "superficial", as opposed to "deep", account of mental states. The first identifies the state with patterns of superficial (macroscopically observable) phenomena while the second identifies it with some properties that underlie such phenomena (Schwitzgebel 2013). This attitude is

---

<sup>119</sup> These are subtle and tricky issues, of course, but since the question of causality is a huge topic on its own right, I won't delve deeper.

<sup>120</sup> See also Alvarez (2010: 49).

contrary in spirit to the essentialist intuition that psychological kinds have essences that need to be captured by scientific investigation. I have also rejected this intuition. Having beliefs is not a matter of instantiating a specific neural structure or even being characterizable in terms of a functional profile that can be defined fully. It is a matter of having dispositions which have open-ended manifestation conditions. As Schwitzgebel puts it:

To believe that *P*, on the view I am proposing, is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that *P*. What respects and degrees of match are to count as “appropriate” will vary contextually and so must be left to the ascriber’s judgement. (Schwitzgebel 2002: 253)

This contextual variation doesn’t entail that attitudes somehow ontologically depend on interpretation. MCC is easily comparable to Schwitzgebel’s conception of beliefs but it also adds to that an account of desires which connects them closely with hedonic dispositions. In any case, the idea that beliefs and desires have their characteristic dispositional profiles, which also reveal their “superficial” nature, can be taken to be a claim about the dispositional stereotypes of those attitudes.

It is important to stress the modesty of the present dispositionalist proposal which distinguishes it from reductive analyses which claim to state the necessary and sufficient conditions for having an attitude (Schwitzgebel 2002: 258). What I take from Schwitzgebel is the idea that dispositions don’t enter into an account of attitudes as defining properties but as elements in a stereotype. Whether or not an having certain dispositions amounts to having a particular belief or desire doesn’t always have a completely determinate answer. By rejecting the aim of finding necessary and sufficient conditions for having attitudes, the present proposal avoids the standard argument against dispositionalism, namely, that dispositional specifications can’t fully exhaust what it means to have an attitude due to other possible attitudes one might have. A fully exhaustive dispositional analysis of attitudes is not a feasible option here. However, this doesn’t mean that we need to go further from dispositions in order to know what beliefs and desires are.

Adopting Schwitzgebel’s dispositionalist outlook also helps us better understand how attitude attributions and attitudes themselves relate to each other. The fact that an attribution needs to take into account behavioural dispositions in case of beliefs and hedonic dispositions in case of desires doesn’t mean that the nature of beliefs and desires is exhausted by these dispositions. These dispositions are only the central components of the stereotypes that characterize the attitudes. Having a disposition to feel pleasure with regard to *p* is neither necessary nor sufficient for having a desire that *p*, there can be defeating conditions under which a person still doesn’t count as having the belief that *p*. The same applies to beliefs and behavioural dispositions.

It can be objected that reference-fixing by a stereotype necessarily leaves the nature of the referent *too* open and undecided. A classic example would be

water the stereotype of which is and has been the transparent, drinkable liquid that can be found in lakes and rivers. Nevertheless, its essence of having the chemical structure of H<sub>2</sub>O was discovered relatively late in human history. This relation between the stereotypical/prototypical properties of a kind and its nature has been exploited by externalists about the mental and is an assumption that is quite ingrained in contemporary philosophy of mind. How are beliefs and desires different and why should we be satisfied with dispositional stereotypes as the final word on the matter?

Schwitzgebel's own answer seems to be that his approach helps us to account for in-between cases where a person instantiates a stereotypical dispositional profile only partially (see Schwitzgebel 2001). In such cases it isn't evident that the subject really has or lacks the belief. This makes sense if having the attitude consists of having certain dispositions which can be possessed only partially. Schwitzgebel adds that the identification of beliefs with stereotypical dispositions explains how the attribution of attitudes can be used as a normative tool to motivate people to conform to stereotypes (Schwitzgebel 2002: 262; 2010: 546).<sup>121</sup> If we treat the belief's stereotype only as a guide to belief's deeper nature and not as constitutive of that nature itself then shaping one's dispositions won't be sufficient for living up to the normative requirement. But that seems strange. Why wouldn't conforming to the disposition be sufficient for fulfilling what the normative attitude-attributions require?<sup>122</sup>

I think that the second answer is actually more promising than the first, although it might seem merely pragmatic by nature. The appeal to in-between cases may easily backfire. One might take the ambiguity of these cases to show that the everyday concept of belief isn't crisp enough and needs to be specified in a way that avoids the possibility of in-between cases. The second response, on the other hand, can help us to see how a concept of such an attitude is actually important. The contentious assumption here is that the ontology of the mental can be informed by pragmatic considerations but this has been the framework of this thesis anyway. In any case, from Dennett and Schwitzgebel we can acquire enough resources to argue that there are interpretation-independent facts about beliefs and desires, without having to commit ourselves to the view that those attitudes have a deeper nature to be discovered.<sup>123</sup>

---

<sup>121</sup> Schwitzgebel isn't explicitly using the normativity-considerations as an argument for the shallow ontology. The following is a quite contentious interpretation.

<sup>122</sup> Since dispositional properties are not usually taken as proper relata of causal relations, Schwitzgebel's position excludes mental properties from having causal efficacy. As already noted I am willing to endorse this position.

<sup>123</sup> I'm thus in full agreement with Schwitzgebel when he says: "The metaphysically inclined may wish to ask whether, useful or not, the present account accurately describes what belief really is. I must admit that I fail to feel the impulse that drives questions such as this—and corresponding questions in other areas of philosophy, such as, What is a person, a cause, or free will, really? There are useful and less useful ways to think about such topics, ways that correspond better with divisions and tendencies of the sort reported by empirical sciences, ways that promote or hinder a particular vision of human flourishing and the development of



### 6.6.6 Affectability and ontology

There still remains a question that needs to be responded to. Remember that attributions of beliefs and desires were analysed in terms of vicarious utterances and attributers' responses to them. Here, in the chapter about ontology, the account I have given has been basically a dispositionalist one. The question is: if I buy into the dispositional account of beliefs and desires, what ontological relevance, if any, is left for affectability and the communicative role of folk psychology? These aspects of beliefs and desires primarily concerned their attribution, not their nature. Yet one might argue that only by ignoring the affectability of folk psychological attributions can I claim that my account of beliefs and desires is a form of realism. After all, if affectability played a part in determining the nature of mental states the facts about ontology would be dependent upon the reactions of the attributers. This would seemingly entail anti-realism about the mental because the truth of mental state attributions would be interpreter-dependent. This isn't a fatal problem but it seems to bifurcate the ontological and epistemological (or attributional) questions in a way that renders the facts about affectability entirely irrelevant for what the ontology of beliefs and desires is.

We seem to have a dilemma, then. Either defend realism about beliefs and desires and pay the price of excluding affectability from the ontological game or claim that affectability has ontological relevance in which case anti-realism seems to follow. I admit that the ontological irrelevance of the communicative significance of mental state attributions is a consequence which may be hard to stomach, considering how important it was for understanding folk psychology.

The issue here might not be that black-and-white, though. The notion of 'ontological relevance' isn't completely clear. The modified communicative conception can be taken to identify the paradigmatic uses of the concepts of belief and desire. In this case the assertability and requestability on others' behalf and the pertinent responses of agreement/disagreement and endorsement/disapproval play a part in fixing the reference of the concepts of belief and desire. They would thus provide us with insight about the dispositional stereotypes that are associated with these attitudes.<sup>124</sup> So affectability is ontologically relevant at least in that sense. This is still a relatively limited sense of relevance because it doesn't entail that having the attitudes of belief and desire requires others to have or be able to have engaged responses to them. But it isn't that

---

moral community. A philosophical account or concept may prove useful in one context or relative to one set of goals and a hindrance in another context or with other goals." (Schwitzgebel 2002: 270)

<sup>124</sup> The stereotypical disposition can be taken as an anchoring instance of the concept (see Kriegel 2011: 14). Unlike Kriegel, though, I don't think that intentional states form a natural kind. Commonality across the instances of belief and desire doesn't derive from the underlying nature but from the features of the stereotype (or anchoring instance, if one prefers), the possibility which also Kriegel actually allows (Ibid. 55), although he doesn't give it much attention.

surprising, considering that the notion of affectability was meant to capture first and foremost the facts about the attribution of those attitudes.

Somewhat more speculatively, in some cases the pragmatic factors of an attributional context may defeat the ascription that is called for by facts about dispositions. In other words, there may be cases when an assertion or request on a person's behalf is not appropriate and thus an attitude ascription isn't true, despite the other person having all the pertinent dispositions. This is in harmony with the claim that having the pertinent dispositions shouldn't be taken to be the necessary and sufficient condition for having the respective attitude. As I've said, it is quite a speculative suggestion because I don't have any clear example nor argument in its support. Nevertheless, it should create some space for a proponent of the communicative conception to argue that facts about beliefs and desires are not completely determined by facts about behavioural and hedonic dispositions, but also by pragmatic factors. It also explains why I don't want to *identify* attitudes with disposition-clusters. Such an identification would preclude the possibility that facts about dispositions don't always determine facts about attitudes.

### **6.6.7 Ontological quietism?**

The final question I want to address in this chapter is whether the present account amounts to a dubious form of ontological quietism. I rejected pure agnosticism about the nature of beliefs and desires. One must accept the minimal assumptions according to which belief and desire attributions need to take into account behavioural and hedonic dispositions and these assumptions have to be acknowledged as capturing something about the nature of those attitudes. But it seems that when we accept that the instantiation of stereotypical dispositional profiles suffices for persons to count as having beliefs and desires, it is impossible to argue against the idea that beliefs and desires exist or that belief and desire ascriptions are true. This seems to entail a strong form of quietism according to which one simply has to take our everyday attributional practice for granted and to be in no need of repair. But surely, the objection might go, it is possible that advances in scientific psychology show that folk psychology is naïve and mistaken. One may feel that our shrugging off of ontological worries has been too hasty.

To this I can only respond that merely feeling that folk psychology has more substantial commitments doesn't amount to an argument. That feeling probably stems from the intuitions about the causal efficacy of the mental, modelled on the relation between physical occurrences. I have already rejected this intuition. Nevertheless, I wouldn't say that my position is a form of absolute quietism because it allows for different ways in which folk psychological attributions can be mistaken. First, there is the quite trivial fact that those attributions can be sometimes false. Since quietism, in order to be contentful, should be considered as a more global thesis about folk psychological attributions the possibility that particular attributions are false doesn't really pose a challenge. But there are further possibilities. For instance, there are cases when a scientific discovery

can have some undermining effect on a whole class of attributions that we make. Among them can even be those that are considered by competent ascribers to be adequate. For instance, the well-known confabulation data from psychological research indicates that people are prone to rationalize their actions by making up mental states that supposedly motivated them but which they actually didn't have (Gazzaniga 1985; Hirstein 2005). This seems to count as evidence against folk psychological self-attributions that are taken to be true also by competent ascribers. It seems, then, that folk psychology is not entirely impervious to scientific data.

But notice that even if the confabulation data disconfirms many attributions that people make, this doesn't amount to eliminativism as a global thesis. It still presupposes the truth of many mental state attributions. Confabulation is a contrastive notion and requires the background of true attributions. Nonetheless, it speaks against a strong form of quietism which implies that scientific psychology can't really speak against competent folk psychological attributions at all and it demonstrates that the present account isn't completely quietist. Psychological research can affect the commonsense picture of the mind with regard to questions about the extent of ascriptions' reliability and their relation to other cognitive abilities. It doesn't threaten the basic assumption of folk psychology that people have beliefs and desires.

The lesson of all this is that scientific theories challenge commonsense conceptions only when there are continuities between them. One reason why eliminativism failed to dethrone folk psychology was because beliefs and desires belonged to the level of coarse-grained dispositions. The motivator for eliminativism, on the other hand, came from considerations about the neural level. If they belong to different explanatory levels, there is hardly any real challenge from one side to another. Those strands of psychology that invoke contentful mental states when explaining behaviour, on the other hand, can challenge the commonsense assumptions. They can do that (to some extent) because they explain human beings on the same level of description. There can be tensions between the two approaches. Something goes wrong when the threat to folk psychology is seen to come from anti-representational accounts of the mind, not from classical cognitive science. The threat from the latter, however, doesn't threaten the elimination of folk psychology. The "worst" it can do is to challenge some ingrained assumptions about beliefs and desires, not their reality as such. The answer to the question as to whether the present account is quietist, then, isn't a straightforward "yes" or "no".

## **6.7 Conclusion**

As I've noted, the modified communicative conception of beliefs and desires that I have defended is first and foremost an account of the attribution of those attitudes. This has probably become apparent in the present chapter where the investigations of the ontology of mental states led us to a rather modest

position. Although it is a form of realism, it doesn't provide us with any insight into the deeper nature of beliefs and desires, if there even is any. It is most relatable to the dispositionalist approaches to attitudes. Still, during our inquiry we hopefully managed to dispel some of the temptation to provide such an insight in the first place. Belief and desire attributions are responsive to reality and this is all we need.

## 7. THE FIRST-PERSON PERSPECTIVE

### 7.1 Introduction

In the previous chapters I have presented an account of the attribution of beliefs and desires which stresses the affectability of such ascriptions and their connection with assertions and requests. I have so far dealt only with third-personal (and second-personal) attributions of attitudes. The question has been what do we do when we think about what others believe and want. But it's clear that beliefs and desires are also attributed to oneself. If my account doesn't take these into account, it can only do partial justice to the facts about folk psychological understanding. It is within first-person ascriptions that mental predicates have been taken to play a peculiar role that makes them difficult to understand.

Probably the most common way to articulate such a peculiarity has been to point out the epistemic immediacy and trustworthiness of self-ascriptions of mental states. Unlike third-person cases, first-person attribution seems to usually guarantee the knowledge that one has the ascribed mental states. In other words, people seem to have privileged access to their own mental states. Self-ascriptions enjoy a special kind of reliability when we compare them with third-personal attributions. This kind of immediacy and reliability may be explained by invoking some metaphysical relation of acquaintance between the subject and her mental life (as found for instance in Fumerton 2005) or by postulating some cognitive mechanism that allows reliable inferences to be made about one's mental states (Nichols & Stich 2002), for instance. It may seem that the modified communicative conception lacks the means to explain such a first-person authority. Although it defined the attribution of beliefs and desires in terms of assertions and requests, the mere assertion or request on the part of an attributee is not sufficient to grant her the respective belief or desire. Also the relevant dispositional and hedonic profiles have to be instantiated. Since it is far from clear that subjects are authoritative about the latter they seem to lack privileged access. Other subjects might know about their dispositions better than they do. In this chapter I will try to argue that one must and can learn to live with such consequences.

But the current account may face an even more immediate problem which is also related to self-ascriptions. One may suggest that it is phenomenologically implausible that we relate to our beliefs and desires through assertions and requests, be they implicit or overt. For instance, the proponents of cognitive phenomenology, at least those who defend the existence of attitudinal phenomenology, would claim that beliefs and desires can be recognized by the subject by their specific phenomenal feels. But even those who reject cognitive phenomenology might argue that the ascription of an attitude to oneself need not be mediated by any speech acts. Although the modified communicative conception doesn't reduce the ascription of beliefs and desires to assertions and requests, it sees the latter as a necessary component in an act of attribution. This might speak against one's everyday experience.

I address the phenomenological worry first and then move on to consider the question of epistemic privilege. In the course of addressing these issues I compare my account of self-ascription with two potentially opposing conceptions. One is the approach of cognitive phenomenology and the other is the account defended by Peter Carruthers'. In a nutshell, what I want to show is that the modified communicative conception I have defended is actually well placed to allow us to understand the relative authority, but also the fragility, of self-understanding.

## **7.2 The phenomenology of beliefs and desires**

### **7.2.1 Affectability and self-ascriptions**

As is the case with third-person ascriptions, first-person cases don't give us any reason to be confident about the predictive model of folk psychology. For instance, if I say or think "I want some coffee" and "I believe that it can be bought from the shop nearby" do I really predict my behaviour? Predicting one's own behaviour seems to imply that I don't conceive the forces that drive me as my own but in the "coffee-situation" I am not detached from my belief and desire. I identify them as my own. In the present context the question is whether affectability also fails to characterize self-ascriptions. By affectability of mental state attributions I meant the basic responses one can have towards the attributed attitudes: agreement/disagreement in the case of beliefs and endorsement/disapproval in the case of desires. At first glance, it may seem quite absurd to say that when one self-ascribes such attitudes one agrees or disagrees with them and endorses or disapproves of them. The relation to one's own attitudes is much more immediate and a person who was so distant from them would seem to count as abnormal. But instead of seeing this as a problem for my account, we can take the implicit agreement and endorsement to be the constitutive elements in self-ascriptions of beliefs and desires. By "implicit" I mean that the response is automatic and doesn't require conscious consideration by the subject. A person need not take an explicit reactive attitude towards her own attitude as in the third-person cases. By attributing a belief that  $p$  to oneself one already agrees that  $p$  and by self-ascribing a desire that  $p$  one already endorses  $p$ . That is one crucial difference between first-person contexts and those wherein the other person is the attributee. In the second case both the positive and negative responses are open to the attributer. Nonetheless, we can talk about affectability both in the first- and third-person cases.

Since these responses are implicit, one might ask: how does this picture of self-ascriptions actually relate to phenomenology if agreement and endorsement are not experientially apparent. The answer is that from the phenomenological point of view there seems to be an immediacy to self-ascriptions that is absent in the case of third-personal ascriptions. Implicit agreement and endorsement are meant to explain this, even if these responses are not introspectively apparent.

Let's try to elaborate the affectability of self-ascriptions further. The reactions that can follow the implicit agreement with or endorsement of the content of one's own belief or desire mirror the reactions in the case of other-ascriptions. Since implicit agreement or endorsement in the case of self-ascription are positive reactions this mirroring is partial. The divergent attitudes are excluded from the normal cases of self-ascription. For instance, in a normal context one can feel resentment towards another's attitude but in the case of self-ascription resentment would be a mark of abnormality. Also, the adoption of one's own attitude (cf. Ch. 3) isn't possible because this applied only to the divergent attitudes. Nonetheless, some possible consequences of self-ascriptions are the same as those that characterize other-ascriptions – having self-ascribed a belief the subject can seek further evidence for it and express it as a pre-condition for cooperating with other agents. Having self-ascribed a desire the subject can try to satisfy it and express it to others. This again might seem trivial but it's important to keep it in mind because these are the basic consequences of mental attributions.

The explication of mental state ascriptions in terms of assertions and requests is also applicable to self-ascriptions.<sup>125</sup> Let's take the self-ascription of desires first, in the form: "I want that *p*." By making such an utterance a person usually conveys a request to her audience or at least makes it clear that it should be the case that *p*. Gilbert Ryle might have got it right when he pointed out that: "...in its primary employment 'I want...' is not used to convey information but to make a request or demand. It is no more meant as a contribution to general knowledge than 'please'. To respond with 'do you' or 'how do you know?' would be glaringly inappropriate. (Ryle 1949: 183)" One might object that this is only relevant when overt expressions of desire are at issue. But even in the cases when a person makes a first-person attribution without addressing other people it can be taken as having a communicative significance for the subject herself. She implicitly endorses the desire. There need not be any explicit utterances or audience involved. The agent is her own audience.

A first-person attribution, then, carries a communicative function on its sleeve and it is quite clearly characterized by affectability, whether the person being affected is the attributee or the attributer herself. An analogous picture applies to belief-attributions. In their first-person employment they can be taken as assertions to oneself that are to be agreed with by the asserter. If a person says "I believe that *p*" she can be taken as making a claim that *p*, even if she says that silently in her privacy. But unlike in the case of other-ascriptions of belief, the subject implicitly agrees with the content of the assertion. It thus gives the impression of immediacy to the subject.

Taking the agent of ascription to be her own audience might sound strange to many. It is as if the person who performs the assertion that counts as a belief-ascription believes what she says. But as Richard Moran notes: "It has been

---

<sup>125</sup> One should keep in mind that assertions and requests in the present sense can also be uttered in private and without vocalization.

noted by more than one philosopher that the relation of ‘believing someone’ does not have a reflexive form; it is not a relation a person can bear to himself.” (Moran 2005: 21) Doesn’t the present account claim exactly that? I don’t think that “believing what one says” counts as an accurate characterization of the first-person belief-attribution. The assertion itself already should count as a belief-ascription, as long as an implicit agreement is in place. It does not involve forming an additional belief, the content of which is the speech act. The agreement that the self-attribution involves doesn’t amount to such a belief.<sup>126</sup>

It is important to keep in mind that self-ascriptions should be taken as first-level assertions and requests. The self-ascription of a belief that *p* is an assertion that *p* and the self-ascription of a desire that *p* is a demand that *p*. This may not seem right to many. After all, one might have an intuition that there is more to self-ascribing attitudes than making the respective speech act. But I do claim that this is exactly what happens when a reflective individual asks herself “Do I (really) believe that?” and “Do I (really) want that?” In order to answer that question, a person should try to assert that *p* and endorse that *p*, respectively. If she can do this wholeheartedly, this counts as a self-ascription.<sup>127</sup> My proposal is, then, that if she is able to wholeheartedly assert that *p*, she counts as having ascribed herself the corresponding belief and the wholehearted demand that *p* counts as the self-ascription of the desire. At this point I take the notion of “wholeheartedness” for granted but I will address the question about its nature in the next section.

Before going further it may be useful to compare this account of self-ascriptions with Robert Gordon’s notion of ascent routines. When we think about our own beliefs from a deliberative standpoint, we don’t usually seem to look inward for the attitudes we have. At least in normal cases, the way to check whether I believe that *p* is to ask if *p* holds (Evans 1982: 225). This method is also known as the ascent routine (Gordon 1996). In deliberating about what I am to believe about X I must give attention to the facts about X, not myself (see also Moran 1994: 161). A person needs to ascertain whether it is the case that *p* in order to decide whether to believe that *p*. It would be irrational for her to believe that *p* unless *p* holds and she cannot maintain that she should believe that *p* while *p* is false. From the first-person standpoint, thinking about my beliefs includes thinking about what is true. This connects quite nicely with my account because in asserting that *p* and implicitly agreeing with it the subject counts as having ascribed herself the belief that *p*. Self-ascription is thus oriented towards the state of affairs that is represented by the belief.

What about desires? Are they also transparent? Bob Gordon has offered ascent routines for them also (Gordon 2007). As it was in the case of beliefs, one might start from considering the content of the desire that *p*. If she is

---

<sup>126</sup> Someone might object that it is simply a matter of conceptual or metaphysical necessity that if a person agrees that *p*, she forms the belief that *p*. But I am not aware of any argument for that claim.

<sup>127</sup> Note that this doesn’t count as *having* the attributed attitude.



willing to endorse  $p$  she can simply “ascend” to a higher, meta-representational level and make a self-ascription in the form “I want that  $p$ ”. One serious difficulty of this account concerns the nature of the content that is being endorsed. The ascent routines were initially modelled on beliefs in the case of which the attitude towards the content  $p$  is simply holding it to be true. That doesn’t seem to work with desires because in order to want a certain state of affairs one can’t take it to be already so. But the latter would be implied if one takes  $p$  to be true. Wanting something seems to be possible only when the desired state of affairs doesn’t yet obtain. Because of that, the ascent routine for desires should have a different kind of input than the ascent routine for beliefs. In other words, in order to attribute to oneself a desire that  $p$  through the ascent routine one cannot ask “is it the case that  $p$ ?” and then attribute to oneself the desire that  $p$  when the answer to that question is affirmative. This problem hasn’t gone unnoticed. Goldman, for instance, has objected that ascent routines can only be applied to beliefs and thus aren’t theoretically interesting (Goldman 2006: 240).

Gordon is much more optimistic and thinks that ascent routines can be understood as automatic redeployments of first-order utterances, so that they can be applied to every type of propositional attitude (Gordon 2007: 154). In order to do that he distinguishes between attitude-expression and attitude-ascription. The first doesn’t imply knowledge about one’s mental states but simply expresses the attitude in question. In that case, an attitude-expression could be redeployed in higher-order self-attribution. For instance, in the case of hope one can use an expressive utterance “I hope that  $p$ ” as the input for the ascent routines. The ascent routine then produces the corresponding self-attribution “I hope that  $p$ ” (ibid. 157). Although Gordon doesn’t mention desires, the corresponding ascent routine for desires would probably correspond to the previous example.

A serious problem with Gordon’s account (which can be avoided if we adopt mine) is that the input to the ascent routine already involves both the attitude and its content. Yet the original idea of an ascent routine revolved around the possibility of starting from considerations about content and then moving to the level of the attitude ascription. In Gordon’s account, the input stage seems to be already as complicated as the higher-order stage.

I think that it is still possible, at least in the case of desires, to hold on to the idea that the first stage of an ascent routine involves only the content of the attitude that will be ascribed in the second stage. The self-attribution of a belief that  $p$  involved the agreement with the assertion that  $p$  is true. The self-attribution of a desire through an ascent routine, on the other hand, could start from asking if  $p$  *should* be true. This would mean considering whether one can sincerely request that  $p$ . If the answer is affirmative then I can also ascribe to myself the corresponding desire that  $p$ . At least on occasions when a person is consciously deliberating over the question of what she wants, this kind of method seems to be feasible and even natural. I don’t claim to be less speculative here than Gordon but my suggestion at least maintains the initial

idea behind ascent routines; that the attribution starts from the first-level utterance. I also dispense with the problematic notion of a non-descriptive expression of a mental state in an utterance. The first stage of the ascent routine doesn't involve any reference to desire nor does it express it. What's more, I don't want to state that this way of ascribing desires to oneself is the only possible method. It is just meant to account for the peculiarity of the first-person perspective. It doesn't necessarily apply to verbally incompetent children. The self-attribution of a desire by means of a request may be confined to a person who is at a certain level of development.

With this initial picture of the self-ascription of beliefs and desires in place, we can now consider certain difficulties it might face. In the end I aim to show that my account isn't inconsistent with the overall phenomenology of beliefs and desires. I insist on separating questions about phenomenology and epistemology. My attention will turn to the latter only in the second part of this chapter.

### **7.2.2 Alienated attitudes**

It seems that there are occasions of alienation when a subject ascribes to herself a belief or desire but can't agree with or endorse its content. The present account treats these cases as abnormalities. It's important to note that here I am not arguing for the claim that alienated belief or desire is impossible or that alienated belief- or desire-ascriptions are impossible. I'm simply claiming that alienated belief or desire ascriptions are abnormal cases of self-attribution because normal cases involve the agreement or endorsement of the content of the attributed attitude. Still, the possibility of an alienated ascription are something that needs to be explained. I try to show that my model of folk psychology can do that if we add some further requirements for self-ascriptions. This is exactly the point at which wholeheartedness steps onto the stage.

So what are cases of alienated attitude? For instance, David Hunter has argued that it is possible to take oneself to have a belief while judging that one ought not to have it (Hunter 2011). This would be exactly the case when the subject ascribes to herself a belief that  $p$  yet isn't (fully) able to agree with it. Although in my account this case is taken as an abnormality, it is perfectly possible. One normally agrees with the self-ascribed belief but there are cases when the normal self-relation is disrupted and the agreement isn't wholehearted. But what could stop a person from wholly agreeing with the self-ascribed belief? I am not sure if the present account can and needs to answer this question because it is ultimately an empirical question. A different question is whether the inability to disagree with the self-ascribed belief is really a common characteristic of the cases of alienated belief. Hunter's examples form quite a diverse bunch. They include both Descartes' alienation from his belief that "one plus two is three" is a necessary truth (Hunter 2011: 223) and a mother's alienation from her belief that her son is doing drugs because she doesn't want to accept it (ibid. 224). In Descartes's case he did seem to agree that "one plus two is three" is a necessary truth while his alienation came from

his other beliefs. The mother, on the other hand, seems to have difficulties with agreeing that her son does drugs. So it seems that only the second case corresponds to the explanation that I've given for alienated belief-ascriptions. But this is not really a problem. It only shows that the case of alienated belief may not cover all of Hunter's examples. Still it is a phenomenon that merits consideration.

The possibility of alienated desires has been much more widely recognized in the philosophical literature. This is quite unfortunate because I don't think that beliefs and desires can be treated asymmetrically. Most notably, Harry Frankfurt has presented a case for alienated desires in his seminal article "Freedom of the will and the concept of a person" (Frankfurt 1971). There he drew a distinction between desires that are the objects of second-order desires and those which are not. The second possibility can be illustrated with the case of an unwilling addict who is aware of herself wanting to do drugs but doesn't want to have such a desire. Although in that paper Frankfurt didn't rely on the term "alienated desire", his discussion can still be taken as a blueprint. The case of an alienated desire exemplifies a situation wherein a person, although having self-ascribed a desire that *p*, isn't fully able to endorse its content. Frankfurt's cases can be seen as drawing the distinction between wholehearted and alienated self-ascriptions (see also Frankfurt 1992).<sup>128</sup> I propose that the notion of wholeheartedness also explains what is missing in the case of alienated self-ascriptions. Although the agent implicitly agrees with or endorses the content of the assertion or request, she doesn't do that wholeheartedly. She is thus unable to fully recognize the attributed belief or desire as her own.

What exactly constitutes wholeheartedness and non-alienation is not an easy question to answer, of course. Frankfurt himself explicates the difference between the willing and unwilling addict in terms of a second-order desire to want or not to want to take the drug. A person would then not be alienated from her desire if she also wanted to have it. I, on the other hand, am not eager to explicate wholeheartedness in terms of second-order attitudes as Frankfurt does because the machinery of assertions and requests, agreement and endorsement, was meant to explain attitude ascriptions. Explaining the wholeheartedness of those attributions in terms of additional attitudes seems to make the account of the first-person perspective quite bloated. Although I haven't aimed at a reductive account of attitude attributions, for the sake of conceptual economy it is still preferable to account for wholehearted self-ascriptions without postulating any higher-level acts or attitudes.

What's more, the higher-level attitude approach is threatened by regress because it pushes us to ask what is it about the second-level mental state that makes first-level ascription wholehearted? This also requires explanation and if

---

<sup>128</sup> It must be stressed that for Frankfurt, this question is motivated by his attempt to explain what the freedom of the will is supposed to be about, he doesn't see this question in the context of self-ascriptions. This means that using his account in the present context needs some translation.

the latter proceeds through the postulation of a third-level attitude then we are on the road to regress (cf. Anderson 2008). Somehow implicit agreement or endorsement of  $p$  on the first level should already be characterizable as wholehearted or not. The case of an unwilling addict, for instance, should be analysed so that the person in question doesn't wholeheartedly endorse the idea that she should take drugs. Whether she wants to want to take drugs or not isn't crucial in determining whether she has an alienated desire or not. An additional reason to reject the higher-level approach is that wholeheartedness should characterize all normal cases of self-ascription. It isn't realistic to assume that higher-level attitudes are necessary for such cases. Finally, the higher-level account is simply phenomenologically implausible. It just doesn't seem to be the case that we need to invoke second-level attitudes in order to make the whole-hearted first-level ascription "I believe that  $p$ ", or "I want that  $p$ ".

What is needed, then? To get out of this pickle, one should observe Frankfurt's alternative observation that wholeheartedness can be understood in terms of being satisfied with oneself and one's attitudes. Although this still seems like a second-order account, Frankfurt has also offered the suggestion that a sufficient condition for being satisfied is that the person isn't actively interested in changing her attitudes (Frankfurt 1992: 12). Since this defines wholeheartedness only negatively, there doesn't seem to be a second-order element included. Although this might not be a fully satisfactory account of wholeheartedness, I accept it for present purposes because the question of interest is whether my conception of self-ascription can allow cases of alienation.

I actually find this negative characterization of whole-heartedness helpful<sup>129</sup> because cases of whole-hearted self-ascription should be seen as a standard condition of relating to oneself. In relation to this, alienated beliefs and desires are abnormalities. It is exactly the latter that demand an explanation and positive characterization. If the agent's self-ascriptions proceed smoothly and involve already implicit agreement and endorsement there is no question of alienation. If, on the other hand, the subject starts to question this implicit self-relation reflectively then agreement and endorsement are still effective but open to explicit questioning. In a way, then, any amount of explicit and conscious reflection about one's beliefs and desires might already bring about an alienated self-relation.

So the overall picture is this: the self-ascription of beliefs and desires proceeds through an implicit agreement and endorsement of the relevant speech act. If the attributer doesn't form any additional attitude of dissatisfaction with these implicit responses then the self-ascription is wholehearted. One becomes alienated from one's attitudes if one does form an additional attitude of disagreement or disapproval. The self-ascription only requires implicit agreement or endorsement. Although wholeheartedness in Frankfurt's sense is explicit, it is more reasonable to see it as a default condition which goes awry

---

<sup>129</sup> As Moran notes: "...wholeheartedness or trust are themselves tacitly defined by the absence of certain thoughts". (Moran 2001: 45)

exactly when the subject's relation to her attitudes becomes her explicit concern.

An objection might be raised, though, that the process of self-ascription I've depicted here does not really concern phenomenology. However this is what is crucial, if we are to understand the first-person perspective on attitudes. This objection forms a natural bridge to the issue of cognitive phenomenology which I will handle in the next section.

### 7.2.3 Cognitive phenomenology?

An account of mental states and their attribution which starts with facts that can be approached from the third-person perspective (i.e. the causal-functional relations between those states and behaviour; the criteria for attributing mental states, such as normative requirements of rationality, or the acts of assertion and request) can face an immediate objection. One might argue that by focusing on what is publicly available such accounts overlook an essential aspect of the mind: its subjectivity. Surely, the objection might go, it is exactly the qualitative and intrinsic character of mental states that presents the most important descriptive and explanatory challenge for theorists who want to understand the nature of mental states. The present account has also been third-personal in this respect and might face a similar objection.

Approaches to intentional attitudes such as belief and desire have faced less pressure to account for their qualitative aspect. It is quite common to claim that attitudes, unlike sensations, are *not* individuated by their phenomenal character. Nevertheless, especially in recent years, the idea of attitudes having characteristic phenomenal properties has gained popularity among those who defend the existence of cognitive phenomenology (see Bayne & Montague 2012). In the present case, desires would also be brought under this label, at least as long as we take them to be propositional attitudes. If the proponents of cognitive phenomenology are right then the account I have defended is fatally deficient. It doesn't seem that I have made any reference to the phenomenal character of attitudes. Isn't there a characteristic feeling of what it is like to believe or want something which is over and above the facts about speech acts and engaged responses? For instance, aren't approval and endorsement already individuated in terms of what it feels like to approve and endorse?

In order to gain some foothold in the whole debate, one should clarify what the thesis of cognitive phenomenology actually amounts to. First, there's the question of what kinds of cognitive state are we talking about if those states are supposed to be endowed with such phenomenology. For instance, Klausen (2008) concentrates on propositional perception and occurrent acts of thinking, and argues that both the content of those acts and the mode by which they are given have a phenomenal character. Chudnoff (2010) has even proposed that there is something that it's like to understand entailment relations. In the present context the question about cognitive phenomenology doesn't really concern beliefs and desires as such. I haven't construed them as occurrent states and the

debate over cognitive phenomenology doesn't deal with the dispositional states (Horgan 2012: 57). Since phenomenal properties are supposed to be introspectively accessible and dispositional states are not like that, it is natural that the latter aren't usually endowed with phenomenal properties. Because of this, the salient question is whether acts of agreement and endorsement have a phenomenal character. They seem to be acts that are present to conscious subjects who seem to have immediate access to them.

Another relevant question is: what is meant by phenomenology? From the perspective of everyday language, it seems like that no one would deny that people experience their beliefs and desires. It is difficult to cut believing or wanting something off from the overall flow of experience people have. This experience is described in relational terms which involve the various ways in which one responds to events in one's environment and these descriptions can be felt as natural and unproblematic.<sup>130</sup> In this regard cognitive phenomenology would be a trivial position and also my description of self-ascriptions would be a description of cognitive phenomenology. After all, I've explained how self-ascriptions of attitudes involve wholehearted agreement or endorsement and how a lack of it makes people feel alienated from their attitudes. Relational descriptions are not what the proponents of cognitive phenomenology have in mind, though. What seems to be at least implicitly presupposed is that the phenomenal character that the cognitive states are supposed to be endowed with is intrinsic and qualitative. It is *not* characterizable in relational terms.<sup>131</sup> Only when this presupposition is accepted is it possible to make the substantive claim that the relational description of attitudes is not sufficient to fully characterize what constitutes them. Deniers of cognitive phenomenology, on the other hand, can be identified with those who deny that the attitudes have intrinsic phenomenal qualities. For those like me, who hope for a deflationary account of phenomenal character which identifies it with relational properties, entering the whole cognitive phenomenology debate carries commitments that are hard to stomach.

---

<sup>130</sup> A lovely description of experience from a perspective that is by no means friendly to phenomenology can be found from Dennett: "What impresses me about my own consciousness, as I know it so intimately, is my delight in some features and dismay over others, my distraction and concentration, my unnamable sinking feelings of foreboding and my blithe disregard of some perceptual details, my obsessions and oversights, my ability to conjure up fantasies, my inability to hold more than a few items in consciousness at a time, my ability to be moved to tears by a vivid recollection of the death of a loved one, my inability to catch myself in the act of framing the words I sometimes say to myself, and so forth." (Dennett 1996: 5)

<sup>131</sup> The cognitive phenomenology has also been defined in terms of „what it's likeness" (Bayne & Montague 2012: 8). But this is a property that can be understood in different ways (Lycan 2008), and I don't see why a functional-dispositional analysis wouldn't suffice to account for what it's like to believe or desire something. At least it's an open question.

I don't think it's possible to settle here the issue whether phenomenal character can or can't be reduced to relational properties.<sup>132</sup> Nevertheless, the previous remarks illustrate how debatable the whole issue of cognitive phenomenology is. As a further example, consider David Pitt's suggestion of a general way of convincing a skeptic about the irreducible phenomenal character of some types of mental state, using sensory states as an example:

A natural response to the antirealist about sensory phenomenology is to induce in him a particularly vivid experience (say, a sharp pain), and then to point out that what he was just aware of is what you claim exists. That is, you acquaint him with the phenomenon you wish to defend the existence of, and argue that the thing just experienced exists. This is admittedly a somewhat crude form of argumentation (if in fact it is really argumentation at all). But I think it does have a significant effect on the polemical situation: it forces the skeptic to deny the obvious. (Pitt 2004: 26).

I find this move on Pitt's part quite revealing because it shows that there might be something wrong with the whole intuition that phenomenal character is a special property of experience. The question one should ask is why Pitt stresses the vividness of an experience as an especially good indicator of the phenomenal character of sensory states. Why doesn't pain *simpliciter* already indicate that phenomenal character is being instantiated? It seems as if the experientiality of mental states (which, after all, no-one denies, at least in some vague sense) should make a more convincing case for skepticism about the intrinsic phenomenal character when the experience is intense. But one should suspect that if it really "forces the skeptic to deny the obvious" then this is merely an nonrational force? Pitt makes a rhetorical move: "You deny the existence of a phenomenal properties but you surely can't deny that you feel intense pain?" When a skeptic hears it, she is put in an uncomfortable position. It may seem that the denial that pain has intrinsic phenomenal properties also implies the denial that one feels pain. But surely this inference isn't valid. For the debate over cognitive phenomenology to be impartial one should avoid simply pointing to a vivid experiential character of beliefs and desires as a way of arguing for their intrinsic qualitative properties.

Let's consider a further example that illustrates why the debate about cognitive phenomenology carries uncomfortable commitments. I have in mind here Christopher Shields's parity argument for cognitive phenomenology. It exploits the following assumption: if there is a reason to endorse the phenomenal character of certain kinds of mental state, such as sensations, then it's equally reasonable to endorse it in the case of cognitive states (Shields 2012: 217). For instance, sensations seem to have determinate intensities which we can ascribe to their qualitative-phenomenal properties. Since cognitive states can also have

---

<sup>132</sup> I agree with Robert Wilson that "phenomenology itself is a contestable phenomenon, and what one can and can't imagine about it inherits that contestability" (Wilson 2003: 427). Because of that, it is hard to say anything conclusive on the matter.

various intensities, it is only reasonable, for Shields, to also endow cognitive states with qualitative properties (ibid. 231). The relevance of the parity argument doesn't have an univocal interpretation, though. If we understand the thesis of cognitive phenomenology in the trivial, relational sense then the parity argument supports the view that believing and wanting, even when construed dispositionally, have an experiential dimension. But this is also quite uninformative because the experiential dimension is assumed from the get-go. If the argument is supposed to strengthen the case for the claim that cognitive states have intrinsic, qualitative properties then this argument is rather dubious. First, it isn't clear that even sensory or perceptual states have such properties.<sup>133</sup> Second, even if we take sensory and perceptual states to have such properties there are further considerations that speak against ascribing those properties to attitudes. For instance, one can rely on an understanding of human cognitive architecture (Carruthers 2009). One can also cast doubt on the putative introspective evidence. Finally, one can do both (Prinz 2012b, Ch. 5).

My strategy thus far has been to cast doubt on the claim that my account of the self-ascription of belief and desire needs to have a response to the question about cognitive phenomenology. If one is a deflationist about phenomenal character then it is quite unproblematic to say that people experience themselves having beliefs and desires and agreeing with and approving their contents. But that is a harmless claim that doesn't commit us to the existence of intrinsic phenomenal properties. The feel of beliefs and desires amounts to the fact that the ascription of them involves agreement and endorsement, respectively, but these acts can be understood relationally. By articulating the self-ascription of beliefs and desires in terms of implicit agreement and endorsement I have already characterized the first-person phenomenology.

A relatively well-articulated argument for cognitive phenomenology is the argument from self-knowledge: since we seem to enjoy privileged access to our propositional attitudes the knowledge of one's own attitudes is immediate. The only way to explain this immediacy is by supposing that the phenomenal character of those attitudes is present to us. Again, Pitt (2004) has made such an argument, for instance.<sup>134</sup> This might be more relevant for the present account of beliefs and desires because I haven't yet explained how we gain knowledge about our own attitudes. If an acceptable explanation needs to appeal to immediate access to one's mental states, there might still be a need to endorse intrinsic qualitative properties that ground such a knowledge. But since the argument from self-knowledge is about the epistemology of first-person ascriptions, I will deal with it in the later section about the epistemology of self-ascriptions.

---

<sup>133</sup> Actually, Shields himself acknowledges this because he takes his arguments to convince only those who believe in qualia.

<sup>134</sup> See Levine (2012: 107), who responds to Pitt's argument by trying to explain the intuition of immediacy with reliable monitoring mechanisms.



#### 7.2.4 Desires and the apparent good

Before moving on to provide an epistemology of self-ascriptions, we should note a certain strand in theories of desire that deserves to be brought out and responded to. Some contemporary authors think that having a desire that  $p$  involves having a cognitive attitude that the object of desire is good (Johnston 2001; Oddie 2005; Scanlon 1998).<sup>135</sup> This idea has a rather long history because it can already be seen in the work of Aristotle:

The object of desire (ὄρεκτόν) always moves, but this is either the good or the apparent good. (*de Anima*, 433a27–29)

Among modern authors, probably the most common motivation for the view that desires relate to the apparent good is that the mere dispositional construal of desires doesn't provide a phenomenologically satisfactory account of how desires reveal certain aspects of the world to us. Desires aren't merely brute causal processes. They seem to present something as worthy of being pursued (Schapiro 2009: 239). Another, related consideration is that without presenting states of affairs in that way, desires couldn't rationalize actions. For instance, Warren Quinn presents a thought experiment in which a person is in a functional state which causes her to turn on radios whenever they are in the vicinity and are turned off (Quinn 1995). Even if this functional state is combined with certain beliefs about how to turn on radios, it still doesn't seem to constitute a desire to turn on radios. According to the apparent good view, what is missing is that the state should present turning on radios as a worthwhile endeavour.

Although this particular thought experiment might be somewhat dubious, I share with the apparent-good theorists the view that in order to understand the first-person perspective to one's own desires it isn't sufficient to characterize desires merely in terms of behavioural dispositions. But the modified communicative conception of desires isn't committed to the view that desires are such dispositions anyway.

If the apparent good view is correct and if we retain the present account of the first-person perspective then the self-ascription of desires should involve a readiness to assert that  $p$  is also good. This readiness explains why the apparent good view is attractive because endorsing  $p$  seems to indicate assent to the claim that  $p$  is good. The self-ascription of desires quite naturally, although not inevitably, involves responses that are pertinent to belief-ascriptions. This hopefully becomes clearer when we consider that agreement and endorsement are modeled on assertions and requests. At least in normal conditions wherein a person requests that  $p$  she also is disposed to assert that  $p$  is, *prima facie*, good. I don't think that this act *necessarily* accompanies self-ascriptions of desire, though. This would be a too strong claim. There are cases when a person isn't

---

<sup>135</sup> This attitude need not be belief. Oddie, for example, is explicit that to desire something is to experience it having value (Oddie 2005: 46).

ready to assert that the content of the self-ascribed desire is good. Just take the case of an unwilling addict, for instance.

However, doesn't the apparent good view exclude infants and animals from having desires? Note that I am not discussing that view in relation to the ontology of desires here. I am only stating that the self-ascription of desires seems to involve the appearance of their content being good. This probably excludes animals from self-ascribing desires but allows the ascribability of desires to them. By operating on the level of attribution, not ontology, I need to interpret the apparent good view in such a way that it could be applicable on that level. This means that I am probably moving away from its proponents' original intentions anyway. A more interesting objection to the apparent good view is that it excludes nihilists from self-ascribing desires. A nihilist thinks that every evaluative judgement is false (Friedrich 2012: 292). The view that self-ascribing desires that  $p$  involves the implicit agreement that  $p$  is good seems to exclude the possibility of nihilists who think that nothing can be judged to be good. But that seems absurd. On the present account, though, this objection is also avoidable. An assertion that  $p$  is good isn't a *necessary* response to a desire-attribution.

One might still doubt, though, if this explanation of the attractiveness of the apparent good view goes deep enough. I have given assertions concerning  $p$ 's goodness only a contingent status. This assertion isn't really a necessary response. But the apparent good view pretends to be a constitutive account or so it at least seems. So let us consider another way of making sense of the intuition behind this view.

We should start with the fact that Friedrich himself wants to explain the apparent good intuition not in terms of belief that  $p$  is good but in terms of what he calls "mental force". He models this on the force of sentences. The content of desire is given in an "evaluative manner" which, for Friedrich should explain why desires involve the appearance of the object being good.<sup>136</sup> By simply stipulating it, this characterisation is quite uninformative. One shouldn't be satisfied with the explanation of the appearance of the good which simply states that the object of desire is presented in an evaluative manner. But Friedrich's proposal is relatable to the account given here. Implicitly endorsing the content that  $p$  in the case of desire ascriptions can explicate what is meant by mental force. To be fair to Friedrich, he himself wants to understand mental force in terms of phenomenal character (Friedrich 2012: 299). We can reject this but still insist that the notion of mental force can be put to work. Perhaps, then, it is unnecessary to suppose that one normally has a belief-like response to the content of one's desire and it suffices to point out that endorsing  $p$  already creates the appearance of  $p$ 's goodness.

I leave open the question of which response is more satisfactory because I don't have any direct argument that would force us to prefer one view over the other. Nevertheless, the second proposal has the advantage of relating the

---

<sup>136</sup> Friedrich himself calls it "the alluringness" of desires.

phenomenology of wanting to expected hedonic states. According to my theory the main consideration that has to be taken into account in the case of desire-attributions is whether the request made tracks the hedonic states of the subject. One way to conceive of the mental force of desires is to take it as the expectation of pleasure in relation to the state of affairs that is represented. One may object that the appearance of the good shouldn't be equated with the expectation of pleasure but I won't delve deeper into that question. If we don't buy into a certain form of platonism about pleasure then it is natural to assume that pleasure exemplifies a kind of good.

Although I haven't provided a definite explanation of how we can deal with the appearance of the good thesis, it doesn't pose any grave difficulties to the account of self-ascription that is presented here. Having characterised the phenomenology of belief and desire it is now time to answer question, how subjects gain knowledge about their attitudes and whether this question is a problem for the modified communicative conception?

## **7.3 The epistemology of self-ascriptions**

### **7.3.1 From ascribing to knowing**

Thus far I have considered what it is to self-ascribe an attitude. How these attributions can be true and provide knowledge is a further question. In this section I will look at the epistemology of self-ascriptions of beliefs and desires. I try to clarify whether the considerations about the peculiar way in which these ascriptions are made also confers an epistemic privilege on them. It has been a quite common understanding in philosophy that first-person access to one's own mental states is epistemically more trustworthy than access to the attitudes of others. This asymmetry has been articulated in various ways: from that access being infallible to it simply being subserved by a reliable cognitive mechanism. What's more, there isn't even a consensus on which kinds of mental states the privileged access thesis is meant to cover. Different kinds of mental state need not be characterized by the same epistemology. Our access to our sensations or emotions, for instance, might be quite different from our access to our attitudes (Moran 2001: 10).

I will try to show that my account of beliefs and desires doesn't need to buy into any strong version of the thesis of privileged access. It only merits its weakest version according to which the epistemic asymmetry exists but only as a contingent feature of the human condition. This feature is rooted in the fact that people usually have more information about themselves than others. This was basically Gilbert Ryle's (1949) position and the following discussion can also be taken as a defence of his views. My arguments are limited to beliefs and desires as dispositional states, though. I won't say anything about the possibility of strong privileged access to sensations, for instance.

As already noted, first-person access to one's own mental states is usually taken to be in some sense epistemically more guaranteed than access to the

others' mental states. There isn't really any agreement on how this should be understood. Let's consider three ways of articulating the epistemic asymmetry between the first- and third-person perspectives:<sup>137</sup>

- 1) **Epistemic security** (ES): if a person (sincerely) self-ascribes a mental state M, her ascription is infallible and incorrigible;
- 2) **Self-intimation** (SI): If a person has a mental state M, she is disposed to know that she has that state.
- 3) **Privileged access** (PA): people have an epistemically special sort of access to their own mental states and not to those of others;

Each of these three ideas could be formulated in numerous different ways. Already incorrigibility and infallibility are quite different characterizations of epistemic security. How does the present account of mental state ascriptions relate to them? ES seems problematic. It is far too strong because it is quite common that people are mistaken about their beliefs and desires. This fallibility should be acknowledged whichever view of self-ascriptions one adopts. It might seem that if self-ascriptions require only sincere assertions or requests, they are reliable because they only require the ascriber's linguistic competence. But note that this doesn't entail that the competence with self-ascriptions guarantees their truth.

As for SI, the first problem is that the conditions of having beliefs and desires, assertability and requestability on a person's behalf, can sometimes be accessible to other people without the subject herself being aware of them. This means that situations should be possible wherein a person has a belief or desire but isn't disposed to self-ascribe it. For instance, I might be hungry and want to eat something nutritious but the hunger isn't intense enough for me to be aware of the desire to eat. What's more, even if we allow for this kind of disposition, a person might only count as having self-ascribed an attitude. This might not count as the knowledge that she has it. True self-ascription doesn't entail knowledge. Again, as in the case with ES one can doubt that the connection between having an attitude and knowing that one has it can be so tight as the formulation claims. The least contentious articulation of the epistemic peculiarity of the first-person perspective is PA. It is also the least informative one. However, it is reasonable to take it as our starting-point, without presupposing any stronger view like ES or SI. The question about the self-knowledge of beliefs and desires, then, amounts to asking, what could the privileged access consist in?

Remember that when I articulated the way in which we self-ascribe attitudes I compared my account favorably with Robert Gordon's idea of ascent routines. This went back to Evans's transparency proposal, i.e. that we ascribe beliefs by focusing on their contents. I also defended the possibility of applying ascent routines to desires, although in a somewhat different manner than Gordon did. One could argue that the method of ascent routines that underlies self-

---

<sup>137</sup> The labels have been taken from Bar-On (2009: 54-55) but aren't intended to accurately represent her use of them.

ascriptions might also bear upon the issue of privileged access and self-knowledge. Alex Byrne, for instance, has proposed the following rule for making self-ascriptions of beliefs which is also supposed to have epistemic import:

**BEL** If  $p$ , believe that you believe that  $p$ . (Byrne 2005: 95)

At first glance this seems to be a reliable rule because even if  $p$  is false, it still seems to give a reason to think that one believes that  $p$ . BEL is self-verifying in that sense. After all, assenting to  $p$  seems to involve forming a belief that  $p$ . It also relates well to the present account according to which a wholehearted assertion that  $p$  is sufficient to count as an ascription of belief. Assenting to  $p$  thus gives a prima facie reason also to believe that one believes that  $p$  or to sincerely assert that one believes that  $p$ .

Byrne proposes a similar rule for the ascription of desires. In order to self-ascribe a desire one needs to consider reasons that count in favor of the state of affairs in question.

**DES** If  $\phi$ ing is a desirable option, believe that you want to  $\phi$ . (Byrne 2012)<sup>138</sup>

Byrne considers a possible objection; namely that his account of desires is circular. Knowing what is desirable seems to imply knowledge about what one desires. Byrne rejects this implication. Considering whether a certain option is desirable doesn't require focusing on one's desires. It can also proceed through the weighing of reasons for the option in question. I am not sure how successful Byrne is in fending off this objection but note that the present account of desire-attributions provides a way to see desire-ascriptions in a different light. The process of attribution doesn't require considering whether  $p$  is desirable. It is sufficient to sincerely demand that  $p$  in order to count as having ascribed the desire. This action can also involve the consideration of reasons but there's less temptation to accuse this account of already presupposing the knowledge of one's desires. This picture is still very similar to Byrne's conception because it also suggests that the self-ascriber focuses on the content, not the attitude itself. So it seems that Byrne's account of transparent self-ascriptions is largely in harmony with my account when it comes to the way that the self-ascriptions are made. Nevertheless, it is an open question whether this method explains the strong epistemic asymmetry between the first- and third-person cases. Perhaps somewhat surprisingly I am going to suggest that it doesn't.

Byrne's hope, when he presents his account, is to explain why people are reliably able to know what beliefs and desires they have. Since reliability is a graded notion I even grant that the subject who follows the method of transparency can have some confidence that her self-ascriptions are accurate. But this may be inherited from the fact that she has had more opportunities to reflect about her attitudes and has more information about herself. The main problem is that I take beliefs and desires to be dispositional, not occurrent,

---

<sup>138</sup> Although Byrne takes desires to be relations to action-types, BEL can be translated into a rule about desires as propositional attitudes.

states. This means that sincere assertion and request are not sufficient for the subject to know that she has the self-ascribed attitudes. I should note that Byrne himself doesn't seem to have any problems taking beliefs and desires to be occurrent states. At least he isn't very explicit about it which means that in a way we are talking past each other. But this isn't really a problem because I can use his account simply as an example of a possible way to articulate the epistemic privilege in contrast with my own.

Remember that in order to count as having beliefs and desires one must instantiate the relevant behavioural and affective dispositions. According to the modified communicative conception, mere assertions and requests are not sufficient for that. This requirement shouldn't be confused with the conditions of having self-ascribed those states. Self-ascriptions only required wholehearted assertions and requests by the ascriber.

The relevance of the dispositional basis for the first-person epistemology of beliefs has been often noted. Nicholas Silins claims: "Judging that  $p$  is insufficient for believing that  $p$ , I take it, because believing that  $p$  requires having various dispositions, where judging that  $p$  is insufficient for having those dispositions." (Silins 2012: 308) Although my account of belief attributions has been formulated in terms of assertions, I take it that knowledge that the relevant dispositions are instantiated is required for knowing that one believes that  $p$ . If the sincere assertion contradicts the dispositional belief (like in the case of an implicit atheist) then the method of transparency fails. So unlike Gendler (2008), for instance, I don't think that a true belief-ascription needs to track the endorsement of its content on the part of the subject. Instead, the dispositional profile that is characteristic of the belief can sometimes have a priority over the explicit judgement of the subject in determining whether the latter has the belief or not. If one is mistaken about instantiating the relevant dispositional profile, one is also mistaken about having the belief.

I insist on the claim, then, that there are cases of failed transparency. This applies also to desires. Their epistemology inherits the possible problems that the epistemology of hedonic dispositions has. One might try to argue that people know what they are disposed to enjoy and have reliable knowledge about what they want because of that. This reliability is more fragile than it may seem, though, because of the failures in affective forecasting. When people self-ascribe desires they give expression to what they think will satisfy them. But there is the threat of being misled by the seeming pleurability of certain options. From the results of empirical research on predicting the pleurability of future outcomes, we can see that people make frequent errors in estimating how strong their affective reactions to different scenarios will be (Wilson & Gilbert 2005). I claim that this also has repercussions for the reliability of self-knowledge about desires. An error in affective forecasting of a state of affairs which one thinks that one wants implies that one actually does not want it.

The difficulty of knowing what really satisfies us and gives us pleasure was already noted by Plato. In *Protagoras*, Socrates admits to his opponent that it is difficult to estimate how distant pleasures can actually be more fulfilling when

compared with immediate pleasures because the latter are more salient (*Prot.* 356c5-d4). This means that people are prone to diachronic illusions about pleasure (Moss 2006). Plato (at least as interpreted by Moss) seems to take privileged self-knowledge of desires for granted and only pleasure from their fulfilment seems to be epistemically opaque for him. But I suggest that if one can be mistaken about the pleurability of the outcome of the self-ascribed desire then one can be mistaken also about the desire itself. The fact that one self-ascribes a desire doesn't mean that one has it. I take it, then, that the failures of transparency characterize both beliefs and desires.

Note that cases of failed transparency don't necessarily imply that the transparency method can't be reliable to a moderate extent. Given that the subject usually has more experience with her reactions and dispositions than other people, her self-ascriptions of attitudes might still be relatively more trustworthy than the ascriptions of mental states to her by others. But this seems to be a contingent feature of our epistemic situation, and also Ryle would acknowledge this.<sup>139</sup>

By rejecting the thesis of strong epistemic privilege I am not denying that there isn't any nonepistemic asymmetry between the self- and other-ascriptions. There are notable differences between these two kinds of mental state attributions and I have already laid out these differences in the first part of this chapter. I even grant that self-ascriptions can enjoy a distinctive status in the eyes of others. If one asserts or requests wholeheartedly one is counted as being responsible for living up to one's self-ascription. For instance, Jane Heal (2002) exploits the analogy between self-ascriptions and promises in which case this ability is realized. Unlike Heal, though, I don't think that ability entitles us to endorse any strong epistemic privilege in relation to one's own attitudes. The dispositions one has take a lot of effort to be moulded in accordance with self-attributions and the success of this ability is quite fragile. Heal's proposal only exemplifies the way in which self-knowledge is a matter of achievement. The ability to live up to one's self-attributions doesn't entail any strong epistemic privilege when compared with other-ascriptions.<sup>140</sup> It doesn't confer any epistemically privileged position on the subject because self-attributions can be as fallible and corrigible as other-ascriptions. Nevertheless, the ability to live up to such ascriptions confers on the subject a *peculiar access* to her beliefs and

---

<sup>139</sup> Actually, Byrne himself might be happy with this kind of analysis. In my eyes, it simply shows how epistemologically modest his account is.

<sup>140</sup> There are also some similarities between my account of self-knowledge and the one proposed by José Medina: "Because of its performative aspect, because it is an acknowledgement, a self-ascription is usually regarded as *prima facie* evidence for the subject's commitment and entitlement to the intentional state she claims to have. But of course acknowledgement does not guarantee entitlement" (Medina 2006: 589). Although I haven't explicated my conception in terms of the normative structure of commitments and entitlements, the quoted view relates to mine quite well because it also denies any strong epistemic asymmetry.

desires. The others don't have such an access because they simply lack such a first-personal ability.<sup>141</sup>

But what about a weaker form of epistemic privilege? Thus far I have been quite vague concerning what the denial of epistemic privilege exactly amounts to. A different way to do that would be to articulate the privilege in terms of self-ascriptions having a special kind of justification. The justification-based account has been developed by Declan Smithies, for instance. Smithies postulates a special kind of justification, introspective justification. Only the attribution of those states which are individuated in terms of their phenomenal character are introspectively justified (Smithies 2012: 277). Although beliefs as dispositional states lack phenomenal qualities, the self-ascriptions of them are introspectively justified because beliefs are individuated in terms of dispositions to cause acts of conscious judgement. The self-ascription of such judgements *is* introspectively justified. Introspective justification comes from the fact that states with phenomenal character are supposed to be both the source and the subject matter of introspective justification. If I consciously judge that *p*, I am justified in believing that I've made such a judgement *because* of that judgement itself.

There are many aspects of Smithies's proposal that are unacceptable. The first of them is the endorsement of the notion of phenomenal character which I find ontologically too inflated. The second is Smithies's definition of belief in terms of a disposition to consciously judge and his dismissal of the role of general behavioural dispositions in individuating beliefs. The third is his conception of introspective justification which I find very obscure and unexplanatory. What is it about phenomenal character that it confers so strong a justificatory status to the state that has it? But I am willing to accept that the idea that self-ascriptions have a special kind of justification has some potential. My proposal is that the introspective justification of self-ascriptions is inherited from the fact that, unlike third-person attributions, the first-person attribution doesn't require performing a vicarious speech act on the behalf of someone. The speech act as such is enough. The justification comes from the fact that there is a kind of immediacy to self-ascriptions that the other-ascriptions lack. I take it to be a weak form of epistemic privilege because it only implies differential responsiveness by others towards the subject in the case of self- and other-ascriptions. Others acknowledge that the subject has made up her mind and intends to live up to the self-attribution.<sup>142</sup> But justification doesn't amount to knowledge.

There's also one advantage which my account of justification has over Smithies's. It relates to the notion of justification. Unlike Smithies, I see

---

<sup>141</sup> The difference between privileged and peculiar access is noted by Byrne (2011: 106), for instance.

<sup>142</sup> An objection might be raised that this isn't actually a form of *epistemic* justification at all. I am willing to grant that. The only reason why I take the notion "epistemic" to be applicable here, is that the self-ascriber acknowledges herself as a subject who commits herself to the project of self-knowledge.



justification in intersubjective terms. The paradigmatic cases of some claim or belief being justified belong to the social context wherein the source of justification is the actual or ideal agreement by other people who have evaluated the claim or belief in question. This is simply a remark about ordinary language use. It seems unnatural to say that a belief is justified because facts pertaining to the intrinsic states of the subject. A person justifies her thoughts and actions to others and earns the justificatory status only when the others acknowledge that. In that sense justificatory status seems to be an extrinsic property which is inconsistent with Smithies's account according to which that status is determined by intrinsic phenomenal character. This is certainly a contentious issue but I won't dwell on it further because it isn't crucial in the present context.

Before going on to respond to possible objections to my account, I try to turn one seeming shortcoming of my account to my advantage. The inability to explain (strong) privileged access is usually taken as a theory's weakness. But in the case of attitudes one can argue that the lack of privileged access is actually a positive fact about our self-conception. The fact that we need to take into account and possibly transform our behavioural and affective dispositions in order to know what we really believe and desire lets us see how self-knowledge is a matter of effort and work. As Josep Corbi puts it:

It seems /.../ that a satisfactory account of first-person authority must take into consideration that there is a significant number of cases where a third party may know better than the agent herself whether she is in a certain psychological condition. In other words, there are some cases where self-knowledge must be regarded as an achievement, and not a trivial one. (Corbi 2009: 326)<sup>143</sup>

For this to be possible postulating strong privileged access may even distort the picture of one's self-relation.

In this subchapter I have argued against the claim that there is a strong epistemic asymmetry between self- and other-ascriptions of beliefs and desires. There isn't a separate and especially trustworthy method of self-ascription. The self-ascription of beliefs and desires through sincere assertions and requests requires behavioural and affective dispositions to be in place in order to be true. This means that the use of the method of transparency doesn't really guarantee reliable self-knowledge. The ability to live up to self-ascriptions may provide a form of epistemic privilege for first person attributions of occurrent acts but self-knowledge of beliefs and desires is still a matter of reflective effort.

---

<sup>143</sup> The claim that the self-knowledge of beliefs and desires can be a matter of genuine cognitive accomplishment has also been defended by Krista Lawlor (2008; 2009).

## 7.4 Objections?

### 7.4.1 Two kinds of self-knowledge?

There is a strong tendency to give occurrent assertions or requests pride of place in determining whether the subject has the respective attitude. This kind of prioritization isn't usually expressed in terms of assertions or requests. Rather, in the case of beliefs, judgements are taken to be the central acts which function as guides to subject's beliefs. This was already apparent in Smithies's account. As far as I know there hasn't been any clear suggestion about what the respective mental act for desires is. In any case, requests haven't been suggested to play that role. Nevertheless, taking for granted the present account of self-ascriptions, where dispositions have epistemic priority, we may ask how it should deal with such a tendency.

To this I can only answer that beliefs and desires have been simply construed in the way that reflects the broad epistemic symmetry between self- and other-ascriptions. There can be different ways of understanding what the concepts of belief and desire refer to and I can't condemn those who understand them in different ways. For instance, one can also postulate "decision-based attitudes", the self-ascription of which *is* self-verifying. These could be identified with pro-attitudes (desires) and dispositions to judge (beliefs), which were considered in 5.5.3 and 5.5.4, respectively. In such a case wholehearted assertions and requests do suffice for having the respective attitude. Only pragmatic considerations can decide which kinds of attitude the concepts of belief and desire should pick out. The discussion in the present chapter may provide a reason why we should think that my construal of beliefs and desires tracks attitudes which are pragmatically more interesting. The reason is that knowledge of decision-based attitudes seems to come too cheaply to the subject. As long as we want to construe the self-knowledge of attitudes as an achievement, which provides a reason to commend the subject who possesses it, adopting a "thicker" conception of beliefs and desires is more reasonable.

Admitting the possibility of different kinds of belief and desire might still make one suspicious. Instead of speaking of different attitudes, why not postulate a difference in access? For instance, Richard Moran maintains the univocity of mental concepts but argues for a distinction between two stances towards one's own mental states. One is the deliberative (or non-observational) stance and the other is theoretical (or observational). Instead of invoking something I've called "decision-based attitudes", Moran claims that there is a special kind of non-observational stance to one's attitudes. This involves making up one's mind on the content of the attitude. To the extent that the knowledge about one's own beliefs and desires arises from deliberation, it is up to oneself to decide whether to believe that *p* and desire that *p*. Knowledge of these attitudes doesn't seem to require any kind of observation or perception. First-person access can be non-evidential. This doesn't mean that the decision to believe or desire something isn't based on reasons. Evidence here means

perceptual evidence and this seems to be unnecessary for self-knowledge from the first-person perspective.<sup>144</sup>

The vehicle of self-awareness for Moran is avowal (Moran 2001: 107). An avowal means explicitly endorsing one's belief and expressing commitment to its truth (Moran 1997: 152). This should imply that self-ascriptions through avowals obey the transparency condition (Moran 2001: 101). One cannot coherently believe that *p* and consider *p* to be false which means that once I claim that I believe that *p* I am also committed to its truth. This is also where the responsibility for one's beliefs comes in. While the avowal of a belief differs from attribution based on evidence about one's psychology, Moran still holds that the same type of mental state is tracked both by avowals and attributions (2001: 89). Although he doesn't analyse the avowals of desires, the same point should also apply to them.

Moran's proposal is in agreement with my account when it comes to the question of what is involved in self-attributions. I appealed to the ability to make wholehearted assertions and requests by agreeing with or endorsing their respective contents. These acts can be identified with avowals in Moran's sense.<sup>145</sup> Because of that I can also acknowledge deliberative stance as an element in self-ascriptions. But since I denied that peculiar access implies epistemic privilege I can't really say that the deliberative stance (mere wholehearted assertion or request) provides any epistemically privileged access to beliefs and desires.<sup>146</sup> It can only provide such access to decision-based attitudes, if any. The modified communicative conception, then, simply rejects the idea that the occurrent acts of assertion and request, or avowals, provide any strong epistemic privilege to self-attributers. The deliberative stance can be incorporated into my account, without granting it more than is necessary.

Another, similar objection to my account is what I call "the expressivist objection". According to it the picture of self-knowledge I have presented creates an epistemic distance between the subject and her attitudes. There is the intuition, going back to Wittgenstein, if not earlier, that sincere avowals of attitudes are such that it doesn't make any sense to ask: "You avowed the attitude but do you really have it?" or "You avowed the attitude but how do you know you have it?" As Matthew Boyle puts it, "If confronted with the question how I know, I want to reply 'Well, it's my thought (my belief, my intention, my toothache, etc.)!'" (Boyle 2010: 13) My answer to the expressivist objection is basically the same as the one I gave to the previous objection. I can agree that these questions aren't really appropriate in the case of decision-based attitudes,

---

<sup>144</sup> Insofar as thinking whether to believe that *p* includes thinking about *p* itself, self-knowledge is also based on evidence, often the perceptual sort. The claim is that ascending from "*p*" to "I believe that *p*" doesn't require evidence.

<sup>145</sup> This kind of identification might invoke many objections. Still, the similarities between the two accounts are salient enough for bringing them together.

<sup>146</sup> To do justice to Moran, in his 2001 book he didn't really concern himself much with the epistemological asymmetry between the two stances, so at least the book might be in agreement with my account.

the self-ascription of which *is* self-verifying. But it is entirely appropriate in the case of the attitudes that I have treated here. The intuition behind the expressivist objection is simply wrong.

But how to explain such an intuition? I suggest that the explanation should appeal to the fact that self-ascriptions have a peculiar justification. Remember that this kind of justification doesn't imply any strong epistemic privilege but still plays a role in our social life. It motivates certain attitudes and expectations towards one another. I am willing to speculate that the acknowledgement of introspective justification might motivate people to exaggerate their epistemic privilege. After all, there is a strong phenomenological asymmetry between self- and other-attributions. The first require mere speech acts. The second involve speech acts *on another's behalf*. This may give the misimpression that self-attributions also involve some sort of special epistemic guarantee. But phenomenological dissimilarity doesn't imply epistemic difference.<sup>147</sup>

#### 7.4.2 Desires and satisfaction

For completeness's sake I want to address a possible objection to my account of self-knowledge. It only concerns the attribution of desires. This relates to the assumption that a true desire-attribution requires an accurate judgement that the represented state of affairs will offer subjective satisfaction. The latter is understood in terms of a positive hedonic state. A potentially counterintuitive consequence of such an assumption is that a person can't know whether she has the desire that *p*, unless she also knows that *p* will satisfy her. Let's call this condition the "satisfaction requirement" (SR). Without the latter the truth of the self-attribution would be accidental at best and accidentally true beliefs don't amount to knowledge. Yet it seems that it is entirely acceptable for a person to self-ascribe a desire that *p*. It is also acceptable to say that she knows she has that desire, even if she doesn't know that *p* will offer her satisfaction. It seems, then, that the present account of self-knowledge of desires is too demanding. Those who take this problem seriously might turn their attention to desires as decision-based attitudes, the ascription of which disregards the satisfaction requirement.

In this subsection I try to dispel this worry and argue that the satisfaction requirement is a reasonable demand on desire-ascriptions. It is noteworthy that a similar question became a focus of debate between Russell and Wittgenstein in the early 20th century. In his "The Analysis of Mind", Russell argues that desires announce themselves to the subject by the feeling of discomfort which

---

<sup>147</sup> I must admit, though, that I have left the question of knowing what one asserts or requests, agrees with or endorses, out of the purview. Such occurrent acts have a place in my account and the question of knowing that one has performed these acts is an important aspect of self-ascriptions. It might be that explaining such knowledge requires more theoretical machinery. But since the knowledge of one's beliefs and desires has been the main issue, I won't address that further question in this thesis.

causes her to act in order to terminate that sense of unease (Russell 1921: 32). Russell's view could be interpreted as implying that in order to know what I desire I need to infer what could put an end to the discomfort. This inference should presumably be based on my knowledge about what offers me satisfaction.

Wittgenstein doesn't accept Russell's analysis. For him the absurdity of Russell's view is apparent because it allows the possibility of a desire for X being satisfied by some other state of affairs than X, as long as it puts an end to the discomfort. Such a sense of absurdity is indicated by Wittgenstein's ironic remark: "If I wanted to eat an apple, and someone punched me in the stomach, taking away my appetite, then it was the punch that I originally wanted" (Wittgenstein 1975: 64). Assume that the self-knowledge of a desire for X were only possible when one had the anticipatory knowledge that X would be pleasing. A person could then be mistaken about attributing that desire if X didn't satisfy her or if she found fulfilment in some other state of affairs. Yet the relation between a desire and its content seems to be tighter, as Wittgenstein notes.

I claim, though, that SR is acceptable, although the phenomenology of desire is not as passive as Russell takes it to be. The satisfaction requirement isn't absurd. It applies to cases wherein adult persons reflect on their desires. It is exactly the reflective standpoint towards one's own attitudes that is the issue in the present context. If such a person stands before the question "Do I really want that X?" she has to consider whether the represented state of affairs will satisfy her. If she is familiar with that type of state of affairs she can be more confident in her desire. After all, she is aware of her hedonic responses to that kind of state of affairs. On the other hand, when it's something unfamiliar, she has a reason to doubt that she really has the desire in question.

Take two persons, A and B, both of whom claim that they want to eat ice cream. Person A has never tasted ice cream before, whereas person B has often eaten it and is aware that she likes it. In the first situation A can ask herself whether she really knows that she wants to eat ice cream, despite the fact that she doesn't fulfil SR. It is entirely conceivable that after having been presented with that question, A begins to doubt her desire. In the case of B, on the other hand, the ascription of the desire is much more certain. The familiarity with the hedonic value of the represented state of affairs implies greater certainty concerning one's desire.<sup>148</sup>

It is important to stress that accepting the relevance of SR doesn't mean that a person couldn't ever claim that she desires a state of affairs when she doesn't know whether it satisfies her. It only gives her a reason to doubt her desire report. The acceptability of self-ascriptions may not require knowledge.<sup>149</sup> But

---

<sup>148</sup> There might be other reliable ways to know that the state of affairs satisfies us, of course.

<sup>149</sup> Such a dismissal definitely requires further argument, but in the present context I will simply take for granted that the acceptability of self-ascribing a desire doesn't require knowing that one has the desire.

she can be fully confident only retrospectively after the state of affairs has been realized. Only after A finally gets to eat ice cream and enjoys it, can she be certain that she really wanted it. If she finds ice cream repulsive she can say that she only thought that she wanted to eat ice cream but actually didn't know whether she did. Reflective individuals have here a perfectly reasonable criterion of self-doubt.

The idea that we can be certain of having a desire for an unfamiliar state of affairs only retrospectively may sound unbelievable. Nonetheless, I am ready to accept the counterintuitive consequences of my claims without giving them up. I will conclude with considering and responding to some possible problems with SR.

First, an opponent may point out desires that seem to contradict SR by definition. Take, for instance, a situation wherein A claims to want to take some drug *because* she has never tried any before. Why should she have any doubts about her desire when she seems to be motivated by the fact that the state of affairs in question is unfamiliar to her? But this is actually a very good example with which to illustrate SR and its scope. First, it is important to stress that not conforming to SR doesn't exclude the possibility that A really did want to consume the drug. She wanted it if it turns out that the drug really does offer satisfaction. Still, person A, as long as she is an individual capable of reflection, has a reason to doubt her desire attribution. When she discovers that the drug didn't offer her any satisfaction, she can ask if she really wanted to consume it. The example here, then, doesn't really shake my claims. It plays with the same intuitions that I have already denied.

The second problem with SR is that it is ambiguous as to whether it should apply to pre-existent desires or to the formation of new desires. Is the question one should ask when considering SR whether I *know* that I want that *p* or rather whether I *should* want that *p*? The second question concerns the formation of a new desire, not the issue of whether one has a pre-existing desire. If SR is meant to help answer the second question it loses its epistemic relevance. My account of the epistemology of self-attributions requires the theoretical reading. Fortunately, the purely practical interpretation doesn't capture the contrast between situations wherein SR is fulfilled and where it is not. It leaves a beguiling impression that people could be always able to decide what desires to form: when SR is met, the subject is given the green light for her desire-formation but nothing speaks against her deciding not to. When SR doesn't hold, she has a reason to be careful if she wants to achieve satisfaction. But in both cases the question about one's desires seems to be a purely matter of choice. Epistemically interpreted, SR doesn't presuppose that. In the case when SR is met, the subject can't simply renounce her desire by denying it. Its reality would resist her denial.

In conclusion, we can maintain that self-knowledge of desires requires knowing the relevant hedonic dispositions that determine whether the fulfilment of the attributed desire offers subjective satisfaction or not. One can maintain that even in the case of desires which may seem to be among the least dubitable

elements of the human psyche, self-knowledge is a matter of reflective achievement. In the final subsection of this chapter I will compare my account of self-knowledge with the one defended by Peter Carruthers, to better articulate the level of analysis on which the present project is located.

## **7.5 Comparison: The Interpretivist Sensory-Access account of self-knowledge**

In his impressive book on self-knowledge, *Opacity of Mind*, Carruthers argues for what he calls the Interpretivist Sensory-Access theory of self-knowledge (ISA). According to it our access to our attitudes and thoughts is enabled by the same Theory-of-Mind (ToM) mechanism that makes possible our access to the attitudes of others. Only access to sensory states is immediate and the attribution of attitudes relies on that sensory information. Early in the book, Carruthers formulates the three central assumptions of his theory:

- a) there is a single mental faculty underlying our attributions of propositional attitudes, whether to ourselves or to others;
- (b) this faculty has only sensory access to its domain;
- (c) its access to our attitudes (or rather to most kinds of attitude) is interpretive rather than transparent. (Carruthers 2011: 2)

Carruthers draws from many different sources to argue for his view. First, there are considerations about human cognitive architecture. Carruthers accepts the model of global workspace (Baars 1988) according to which outputs of sensory systems are broadcast across the cortex and thereby made available to different modular conceptual systems which draw domain-specific inferences from those outputs. In the case of self-attributions (and also other-attributions), the relevant conceptual system is the ToM mechanism. The arguments of the second type come from evolutionary considerations. Carruthers argues that it is reasonable to postulate selection pressures for building the ToM mechanism but the evolution of a specific metacognitive mechanism is much less plausible. This indicates that there is probably no separate system which is responsible for self-attributions (Carruthers 2011: 65). Third, evidence from autism, schizophrenia and alexithymia indicates that the ability to make self-attributions isn't dissociated from an understanding of the attitudes of others (Ibid. 293). Fourth,<sup>150</sup> the fact that people are prone to make frequent confabulatory self-attributions when sensory information is misleading seems to demonstrate that access to one's attitudes is interpretive and not direct (Ibid. 325). Carruthers, then, relies on cumulative empirical evidence in arguing for his theory and covers most of the pertinent psychological literature. In the present context I am not going to argue against his general psychological theory. The aim is to explicate the overall differences between his account of self-knowledge and mine.

---

<sup>150</sup> Carruthers' arguments are not limited to these four.

One noticeable similarity between Carruthers' account and mine is that his conception of self-knowledge also rejects the epistemic asymmetry between knowing one's own mental states and those of others. Both ISA and my account state that the difference between self- and other-knowledge boils down to the contingent difference in the quantity of information persons have about their own and others' dispositions. Carruthers doesn't deny that there is more sensory information available to the subject to make inferences about her own attitudes than to others.

There are crucial differences, though. Carruthers mostly focuses on occurrent attitudes. It is true that epistemic privilege towards dispositional states isn't really a position that is often argued for and Carruthers' focus isn't surprising. After all, challenging the claim about epistemic privilege towards occurrent states is a much more potent move. The claim that dispositional attitudes lack privileged access may seem quite unproblematic. Still, given that according to Carruthers access to *all* non-sensory states is interpretive, his account seems to be in harmony with mine. The claim about interpretive access also covers dispositional attitudes. But there's a difference which is even more salient than the similarity. Carruthers' account is almost entirely about subpersonal processes (above all, the functioning of ToM) that underlie self-attributions. My account of the first-person perspective, on the other hand, has been highly intellectualist. I have taken the ability to assert and request to be among the central abilities around which the self-ascriptions of attitudes turn. In addition, my approach has focused on the personal level where the attributions are conscious and reflective. This approach is pretty much opposed to that of Carruthers. But it might still be fruitful to compare his account with my own in order to show that the present conception can add something new to our understanding of self-knowledge.

Carruthers is adamant about insisting that philosophical theories of self-knowledge which claim to be dealing with attitudes on the personal level can be refuted by discoveries in cognitive science. They don't enjoy any independence from the empirical work in psychology (Carruthers 2011: 25). I try to argue, though, that he is too dismissive of the possibility that the philosophical perspective on first-person authority might be (relatively) autonomous from psychological approaches. I must stress that this doesn't mean that we can argue for immediate or privileged access to the attitudes on the personal level. I only want to make the point that the account of self-attributions and self-knowledge that I have developed here can give us insight about the first person perspective on one's attitudes and this isn't dependent upon which theory of cognitive architecture is true.

So why do I think that a philosophical account of self-knowledge can be relatively autonomous<sup>151</sup> with regard to psychological findings? A blunt answer

---

<sup>151</sup> By "relatively autonomous" I mean that it is not necessary for a philosophical account of self-knowledge to be informed by psychological research in order to be feasible. Denying



to this question is that Carruthers and I deal with a different subject matter. We thus have different commitments concerning what a philosophical conception of self-knowledge should achieve. This difference doesn't merely amount to the distinction between occurrent and dispositional attitudes, though, but cuts a bit deeper. To see this we must consider the fourth chapter of his book where he tries to dismantle three different accounts of self-knowledge that have proposed some sort of immediate access to one's attitudes. His commitments become apparent when he discusses constitutivist accounts. He takes the core idea of constitutivism to be that "our own actions (especially in the form of inner or outer speech) can constitute the adoption of a novel attitude via commitment" (Ibid. 96).

Formulated in these terms my account could also be taken as a form of constitutivism. After all, an important element that grounded the peculiarity of self-attributions was the ability of the subject to live up to them. I refused to build any strong form of epistemic privilege on that ability, though. But what's important here is to see how Carruthers deals with the constitutivist accounts. First, he endorses the dual-process reasoning theory and the distinction between System 1 and System 2 processes. The first involves the processing of attitudes which is unconscious and fast. The second is conscious, slow, and reflective (for a general overview, see Evans 2008). This model has also been extended to distinguish between System 1 and System 2 attitudes – the latter being realized by the first (for the most thorough treatment of this distinction see Frankish 2004). Carruthers then argues that even if a constitutivist claims that being committed to an action in accordance with a self-ascription constitutes that attitude on the personal level she can't confine her attention to the System 2 level (Carruthers 2011: 102). There are further conditions for really knowing that one has the attitude, and this requires the System 1 attitudes to be in place. However, access to the latter is in any case interpretive, which seems to mean that the constitutivist can't deliver what she promises.

This argument is telling because it indicates that Carruthers takes the System 1, subpersonal and unconscious attitudes to be the primary objects of self-knowledge. For Carruthers, the subject's exercise of her ability to live up to her attributions isn't sufficient to constitute the attributed attitude.<sup>152</sup> The ToM mechanism needs to make the accurate attributions of the attitudes that realize it. We may safely assume, then, that Carruthers takes the System 1 states as paradigmatic mental states that are the object of self-knowledge. The present account, on the other hand, takes the beliefs and desires as dispositions to be

---

the necessity condition doesn't imply that psychological findings couldn't inform philosophizing.

<sup>152</sup> From the previous discussion, one might have had an impression that I can't admit that either. But note that I was not denying the sufficiency of exercising the ability to constitute the attitude. I was denying that the ability endows the subject with an epistemic privilege. The knowledge of one's dispositions is fragile and the others can know them better than the subject, at least in principle.

central.<sup>153</sup> A person having a belief that *p* is constituted by certain behavioural dispositions that determine whether asserting that *p* on her behalf is appropriate. A person having a desire that *p* is constituted by pertinent hedonic dispositions. The knowledge that one has such dispositions constitutes self-knowledge of belief and desire. These attitudes are clearly different from what Carruthers conceives of as beliefs and desires, namely, representational states “that causally underlie the relevant dispositions (2011: xiv). In any case, instantiating a mere dispositional profile is not sufficient for having an attitude for Carruthers because “*acting as if one* has a given attitude doesn’t entail that one really does have that attitude. (Or at least, not unless one is some sort of anti-realist or instrumentalist about the mental.)” (Carruthers 2011: 102).

*Pace* Carruthers, the modified communicative conception takes dispositions to be sufficient for having an attitude. I do not think that grounding belief and desire attributions in dispositional profiles entails anti-realism or instrumentalism about the mental. I assume the view, which I defended in the previous chapter, that dispositions are as real as their categorical bases. An ascription of beliefs and desires can be true and responsive to reality. Defending this claim further would steer us away from the context of the present chapter, though. What’s important here is that if we accept the conception of attitudes I have presented then my account of self-knowledge can be taken as having a different topic.<sup>154</sup>

So why is the comparison with Carruthers’ theory relevant here, if it is so different from mine? I think it is relevant because it makes explicit the gulf that can exist between different approaches to the question, what does self-knowledge consist in? The gulf is real, although the objects of interest are seemingly the same; namely, beliefs and desires. Unlike Carruthers, I see the attitudes, as construed by the modified communicative conception, to be much more central to our everyday conception of the mind, whether or not this construal is also reflected in our cognitive architecture. There are many situations in our everyday life wherein the questions “What do I believe?” and “What do I want?” matter to us and I suspect that my conception of beliefs and desires is appropriate in such situations. Carruthers himself speaks also about “our ordinary conceptions” when he mentions the attitudes which he centres upon (2011: 110). But he doesn’t really give any arguments for the claim that his understanding of beliefs and desires coincides with common-sense understanding of them.

He does address the possibility of construing the reflectively ascribed attitudes as separate kinds of attitudes but doesn’t think that it is fruitful because

---

<sup>153</sup> Beliefs and desires as dispositions don’t correspond to System 2 attitudes. But they definitely contrast with the attitudes that Carruthers focuses on.

<sup>154</sup> Carruthers briefly considers the suggestion of redefining attitudes, but dismisses this proposal, largely on the grounds that these would be revisionary of the functional profiles of those attitudes, thus losing distinctions between them (Carruthers 2011: 108–113). I, on the other hand, don’t think that different attitudes need to have strictly distinguishable functional profiles in the first place, so I don’t share Carruthers’ fears.

this wouldn't increase the explanatory power of attitude ascriptions (2011: 114). But *contra* Carruthers, in the chapter about affectability I have presented different advantages that the consciously and reflectively applied attitude-concepts can provide. These advantages don't reduce to the explanatory or predictive power and they aren't merely the outgrowth of our intuitions about the transparency of self-attributions. Carruthers might object that he can only rely on theoretical usefulness as a criterion for introducing concepts. But the issue of self-knowledge in philosophy isn't only a theoretical problem of understanding what cognitive mechanisms enable it. It is also a question about our general self-conception, its consequences and its possible improvement which means that the importance of attitude ascriptions isn't limited to theoretical gains.

A more interesting objection by Carruthers is that the reflected attitudes simply aren't very important. In everyday human life one needs to rely on simple heuristics and snap judgements (Carruthers 2011: 116). People often don't act in accordance with the commitments they have undertaken when they self-ascribe attitudes. To that I can only answer that reflective self-attributions are important enough to be accounted for when we bear in mind that we sometimes do think of our attitudes in such a way. My account wasn't meant to cover all aspects of self-understanding anyway. I also acknowledged the fact that people often don't live up to their commitments.

We can say, then, that Carruthers' theory of self-knowledge doesn't exhaust the possibilities of understanding what self-knowledge of attitudes consists in. There is also the personal-level perspective that characterizes us as reflective agents. From that perspective the question of self-knowledge doesn't address the subpersonal functioning of our cognitive system but instead more coarse-grained features of our active and affective lives. One shouldn't let oneself be distracted by the fact that both Carruthers' account and mine deny epistemic privilege to the subject and that both approaches see the first-person perspective in interpretive terms. There is still quite a strong difference between the two. One can't exclude the possibility, though, that the two accounts are actually compatible with each other because they are located on different explanatory levels.

## 7.6 Conclusion

In this chapter I addressed the question how to understand the peculiarities of the first-person perspective to one's beliefs and desires in the framework of the communicative conception. The self-ascriptions differ from the third-personal attributions because they circumvent the need to assert or request on someone's behalf. They are constituted by wholehearted assertions and requests *simpliciter*. The wholeheartedness is defined in terms of the absence of interest in changing one's attitudes. My account doesn't presume that the self-ascriptions have an intrinsic phenomenal character but understands their phenomenology relationally. In addition, the self-ascriptions of desire may involve the

appearance of the good to the ascriber which can be understood in terms of the expectation of pleasure.

As for the epistemology of self-ascriptions, the modified communicative conception need not buy into any strong form of first-person privilege. One can gain knowledge about one's desires by taking into account one's behavioural and hedonic dispositions just as in the case of other-attributions. The epistemic authority of self-ascriptions boils down to the fact that the person usually knows more about her own dispositions than others do.

## 8. CONCLUSION

The primary purpose of this thesis was to articulate the notion of affectability and its implications. Folk psychology is an engaged enterprise in which our understanding of what other people believe and want shapes our responses to them in particular ways. This means that attributed attitudes have communicative significance. Belief attributions enable the responses of agreement and disagreement. Desire attributions, on the other hand, allow for the possibility of endorsing or disapproving of an attributed attitude. The central claim of the thesis is that without such communicative significance the practice of attributing attitudes would be unrecognisable and alien to us. By bringing out the main responses that folk psychological attributions involve, I tried to show why we should care about having the concepts of belief and desire. The reason for caring wasn't confined to the mere fact that there are engaged responses. I also showed how the concepts of belief and desire make it possible to treat the behavioural and hedonic dispositions of other people as if they had communicative significance. This way of understanding beliefs and desires endows their attributers only with minimal ontological commitments according to which having the aforementioned dispositions amounts to having beliefs and desires. I consider this to be an advantage of my account but it blocks us from radically criticising folk psychological discourse. Finally, my account has mostly been third-personal but it can also make sense of the peculiarities, even if not the privilege, of self-ascriptions.

Although I addressed the ontological status of attitudes in this thesis, the focus has been on their attribution. This shouldn't be surprising, given that my approach took inspiration from such philosophers as Wittgenstein and Ryle (and like-minded present day authors) who took the analysis of the use of terms in our forms of life to be the proper way to gain insight about the phenomena which these terms are supposed to designate. This thesis can be taken, among other things, as an attempt to vindicate such an approach. What people believe and want is a question that arises in our linguistic practice wherein we have shared modes of response to one another and this question should be answered in the context of this practice. I allowed for the possibility that folk psychology could be revised or criticised but it turned out that the benefits of applying the concepts of belief and desire don't motivate any substantial ontological commitments. We can maintain that belief-desire psychology is a useful tool for thinking about one another and that there is no pressing reason to assume that this usefulness is grounded in any hidden essences of belief and desire that wait to be discovered. However, this does not mean that there are no aspects of folk psychology which have yet to be properly appreciated. Indeed, until now, affectability has been one of these aspects.

I want now to consider briefly the more general importance of this thesis and to mention some open questions for further research which are motivated by this thesis.

Let's begin with the general significance of this thesis. First, my argument for the indispensability of affectability was meant to say something about the nature of folk psychological interpretation. The indispensability of engaged responses doesn't indicate that the possibilities of agreement/disagreement and endorsement/disapproval are highly frequent, but still merely contingent, effects of mental state attributions. Rather, affectability is comparable to the principle of rationality, invoked by such authors as Davidson, according to which the assumption that attributees are rational is a constitutive principle of folk psychological interpretation. The assumption of affectability (that it should be possible to be affected by one's belief and desire attributions) should also be taken to be such a constitutive principle. As argued in the second chapter, the folk psychological practice as we know it wouldn't exist without affectability. Perhaps it would be possible to argue that affectability actually expresses the assumption of rationality in some way but I am agnostic about that. What matters is that affectability is as important for understanding the nature of folk psychological interpretation as the assumption of rationality is. It makes the practice of folk psychology possible.

One might think that the admission of affectability among the constitutive conditions of mental state attribution spells trouble for scientific study of the mind. After all, cognitive scientists speak of representational mental states without assuming that they have to be able to agree or disagree with and endorse or disapprove of them. Scientific psychology presumably isn't an engaged enterprise (with an exception of psychiatry). Furthermore, my alliance with ordinary language philosophy in the introduction might have left an impression that I want to accuse scientific psychology of distorting our everyday practice of mentalistic interpretation. However, by defending the indispensability of affectability, my aim hasn't been to play a role of conceptual police. I presume that it is legitimate for psychologists to attribute whichever kinds of mental states they do, possibly also to creatures who don't possess language. If they don't respond to attributees in ways which are characteristic of affectability, this only means that they are not engaging in folk psychological practice but are instead doing something different. One should just be heedful of the differences between conceptual practices in order to realise when we are changing the topic. The talk of representational mental states in science can be perfectly legitimate, whether or not affectability is acknowledged there. Nevertheless, from a philosophical standpoint, it is important to keep in mind the differences between everyday and scientific understanding. Affectability is crucial for the first.

The further consequences of affectability which I explored in the third chapter should strengthen the case for affectability's importance. If the attributee is in some way socially superior to the attributer, the latter is prone to adopt the attributed attitude and have a negative emotional reaction towards the attributee. This indicates the extent to which attitudes of other people matter to us and how this is modulated by social relations between attributers and attributees. The attributee's beliefs and desires function as claims and demands,

respectively, and they depend on the relation the attributer reacts to them in different ways.

This brings us to the second important point which can be gathered from this thesis, namely, that folk psychology involves speaking on the behalf of others. This was Gauker's original insight. I retained it in the modified version of the communicative conception. This means that folk psychology has an ethical dimension. When people attribute beliefs and desires to one another, there is always the question: who has the right to speak on whose behalf (i.e. to attribute an attitude)? On certain occasions it may be inappropriate to attribute attitudes to another person. For instance, when an attributer and an addressee are in socially unequal positions, it is possible to argue that, by speaking on the behalf of another person, one is using her superior position to exclude another from speaking on her own behalf. The modified communicative conception hopefully lets us see the ethical tensions in the folk psychological practice more clearly than accounts which don't recognise the communicative significance of mental state attributions.

Another important consequence of this thesis is that it allows for ambiguities in the terms *belief* and *desire*. The modified communicative conception I've defended understands beliefs and desires in terms of behavioural and hedonic dispositions and the attribution of beliefs and desires in terms of assertions and requests on an addressee's behalf. To many it may seem to be a counterintuitive understanding of those attitudes. Although I've fended off the idea that our theory of beliefs and desires has to respect the intuitions of concept-users, I grant that the terms "belief" and "desire" may identify different attitudes on different occasions of use. For instance, I pointed out in the seventh chapter that one can also talk about decision-based attitudes. The terms "belief" and "desire" might not always identify attitudes that I have considered in this thesis. But this is a conclusion that is difficult to avoid if we look at the messy details of the application of these terms. I actually think that recognition of the ambiguities in the terms "belief" and "desire" is necessary if one wants to understand how they are applied in different contexts. Although some may be frustrated by the ambiguity, it is simply the price we have to pay if we want to adequately understand how the mentalistic discourse functions.

What's more, ambiguity makes it natural to consider the historical development and cross-cultural differences in the way in which beliefs and desires have been conceptualized, a possibility that seems to be somewhat overlooked in the contemporary philosophy of mind (although the recent work in experimental philosophy seems to point in that direction). If it is acknowledged that mentalistic talk refers to a plethora of attitudes already within a spatiotemporally limited part of the world, it is easier to accept that there might be noticeable historical and cross-cultural differences as well.

The fourth important take-home message of this thesis is that the ontological commitments of the folk are flexible. It is legitimate to prescribe certain commitments to concept-users or withdraw some from them, even if they find them unintuitive. This became apparent in the sixth chapter where I argued that

by adopting the pragmatic approach to the concepts of belief and desire we can argue that folk psychology should commit one to only minimal assumptions about the nature of beliefs and desires. We need to accept only those commitments which are necessary for concept-application to deliver its standard benefits. Instead of poking at the intuitive judgements of the folk, we can instead look at what they are doing when they use the concept in question. By minimising folk commitments concerning beliefs and desires it was possible to argue for a (nonreductive) dispositionalist view of these attitudes. The latter hasn't been as popular an option among philosophers as a representationalist position, for instance. One of the reasons for this might be the intuition that the folk are committed to more than a mere instantiation of behavioural and hedonic dispositions. From the present perspective, we can now see that it is entirely legitimate to give a minimalist interpretation to folk commitments.

There is one point where the intuitive judgements of the folk might play a role, though. This concerns what people want to achieve when they attribute beliefs and desires to one another. What they are committed to depends on this. I can imagine a possible argument according to which we can't keep the commitments minimal because at least some people want to describe internal states of individuals when they attribute beliefs and desires. On phenomenological grounds, though, I don't think that it is reasonable to claim that in our everyday interactions we (regularly) want to do this. An important thing to note is that a conditional claim remains plausible in any case: one can be an ontological minimalist and dispositionalist about beliefs and desires when one's conceptual needs are modest. The needs that folk psychology expresses *are* relatively modest, as I tried to show in this thesis. The fact that they are modest, though, doesn't mean that they don't merit our philosophical interest.

Before concluding, I'd like to bring out some questions for further research that this thesis can motivate. First, one can and should articulate further the structure of engaged responses which constitute affectability. In my thesis I remained content with a pretheoretic understanding of what agreement/disagreement and endorsement/disapproval consist of. But a closer look might reveal that these responses have quite a complicated structure. For instance, are there differences between having an engaged response to a content-component and having it to an attitude-component, or can one be reduced to another? As I noted in the second chapter, beliefs and desires may require asymmetric treatment in this regard: the response to belief is content-oriented while the response to desire more naturally takes the attitude as its object. In addition, there might arise difficulties with accounting for agreement and disagreement with certain kinds of belief (*de se* beliefs for instance, in which case it isn't clear whether the attributer can entertain the content of the attributee's belief if such a content is essentially self-locating).

A further articulation of the structure of affectability is also needed for responses which are grounded in agreement/disagreement and endorsement/disapproval. I called it affectability in the wide sense. The account of them



presented in this thesis (in the third chapter) has been quite speculative. For instance, the extent to which the adoption-through-attribution model characterises everyday social interactions is open to debate. Also, the analysis of reactive attitudes only brought out the most basic ways in which people can be emotionally affected by belief and desire attributions. In addition, I only hinted at the possibility that the control of others' folk psychological interpretations of oneself may constitute a form of social power over others. Given that the focus of this thesis has been on affectability in the narrow sense, the preliminary character of the investigations into affectability in the wide sense should be understandable. Its full articulation might require a separate monograph.

Second, my thesis can motivate empirical investigations into how engaged responses to attitudes evolved phylogenetically. For the sake of analytic clarity, I have analysed them in terms of speech acts. Although I suspect that their functioning can't be explained without reference to speech acts, it is possible that some sort of co-evolution account would be needed to explain how they appeared in the human lineage. The abilities to comprehend speech acts and to respond to attitudes with agreement/disagreement and endorsement/disapproval might be rooted in some cognitive ability that is itself evolutionarily prior to them. One place to look for such ability is gesturing behaviour. In gesturing behaviour it is possible to distinguish between imperative and declarative pointing. For instance, we can hypothesize that engaged responses that are characteristic of affectability arose in their rudimentary form already in relation to these two types of gesture. More broadly speaking, it is an open question why members of our species began making differential normative evaluations of one another's behaviour which are expressed in the engaged responses that I have analysed. An answer to this question should dissolve the possible worry that affectability, as I've articulated it, lacks naturalistic credentials. Affectability is certainly grounded in our psychological capacities and they have an evolutionary history.

Third, I've relied on the method of conceptual explication (primarily inspired by Justin Fisher's pragmatic conceptual analysis) to articulate the application conditions of the concepts of belief and desire. The explication started with an analysis of the benefits of these concepts and then asked what should be the conditions under which to apply these concepts. Methodological assumptions behind this approach were left somewhat implicit. For instance, although an explication is allowed to be counterintuitive for concept-users, there should be a point at which it can't be said to deal with the same concept anymore. Where exactly does this point lie and is there a principled way to identify it? These are open questions. Also, in this thesis I've stressed that at least the benefits that are supposed to be explained by the explication (affectability, explanations, predictions, mindshaping) should be recognisable to the concept-users. Still, one could inquire further into the question of how to individuate the benefits of a concept and to what extent it is an empirical question, as opposed to an armchair one. To what extent do we have epistemic

authority over what we want to achieve with our concepts and what we actually achieve?

This last question brings us to the fourth topic of research that grows out of the present thesis: the first-person epistemology of beliefs and desires. Although my analysis didn't reveal any strong asymmetry between first- and third-person perspectives, it is an open possibility that people have privileged access to what they assert and request. These are occurrent acts, as opposed to dispositional states, and they are usually considered to be epistemically accessible to their subjects directly. Self-ascriptions of belief and desire were understood in terms of assertions and requests and the epistemology of the latter is yet to be accounted for. I doubt, though, that this has any bearing on the question of whether we have privileged access to our beliefs and desires. What I actually find to be a more promising possibility in this regard is the possible role of imagination in providing self-knowledge. Although I didn't address it in this thesis, one might argue that by imagining a state of affairs which is the object of a self-ascribed desire, and feeling pleasure in doing that, the ascriber has a reason to be more confident in her self-ascription. The desires were understood in terms of dispositions to be pleased. It might be the case that those who feel pleasure in imagination are at least in certain cases justified in assuming that they have a disposition to actually feel pleasure, i.e. they have the desire in question. In a similar way, one may have justification for one's self-ascription of belief if one can imagine the manifestation of relevant behavioural dispositions in response to a state of affairs which is the object of the belief. However, it is an open question whether imaginings can confer any justification on self-ascriptions and the answer seems to depend on the extent to which imaginings can track the actual dispositions that people have.

An additional topic to be investigated further is the explanatory role of beliefs and desires as conceived in this thesis. Since I didn't conceptualize these attitudes as causally efficacious inner states, I had to endorse (in the fifth chapter) an account of context-placing explanations according to which folk psychological explanations serve to articulate more fully the situation in which agents find themselves. This is a controversial thesis and I was admittedly sketchy about it, which means that there is work to do. Finally, I will only mention the historical development of the concepts of belief and desire as an additional topic for further research. I assume that there are other topics which I haven't mentioned but this should suffice to show the significance of affectability and the modified communicative conception.

## REFERENCES

- Allen, Colin. (1992). Mental content. *The British Journal for the Philosophy of Science*, Vol. 43, No. 4, pp. 537–553.
- Alvarez, Maria. (2010). *Kinds of Reasons. An Essay in the Philosophy of Action*. New York: Oxford University Press.
- Amberber, Mengitsu (ed.). (2007). *The Language of Memory in a Crosslinguistic Perspective*. Amsterdam: John Benjamin's Publishing Company.
- Anderson, Joel. (2008). Disputing autonomy. Second-order desires and the dynamics of ascribing autonomy. *Sats – Nordic Journal of Philosophy*, 9, pp. 7–26.
- Andrews, Kristin. (2002). Interpreting autism: critique of Davidson on thought and language. *Philosophical Psychology*, 15, pp. 317–332.
- Andrews, Kristin. (2003). Knowing mental states: the asymmetry of psychological prediction and explanation. In Q. Smith & A. Jokic (Eds.), *Consciousness: New Philosophical Perspectives*, pp. 201–219. Oxford: Oxford University Press.
- Andrews, Kristin. (2009). Understanding norms without Theory of Mind. *Inquiry*, 52, pp. 433–448.
- Andrews, Kristin. (2012). *Do Apes Read Minds? Toward a New Folk Psychology*. Cambridge, MA: MIT Press.
- Andrews, Kristin. and L. Radenovic. (2006). Speaking without interpreting: a reply to Bouma on autism and Davidsonian interpretation. *Philosophical Psychology*, 19, pp. 663–678.
- Anscombe, G.E.M. (2000). *Intention*. Second Edition. Cambridge, MA: Harvard University Press.
- Aristotle. (1941). *De Anima*. In R. McKeon (ed.), *The Basic Works of Aristotle*. New York: Random House.
- Atran, Scott. (2001). The trouble with memes: inference versus imitation in cultural. *Human Nature*, Vol. 12, No. 4, pp. 351–381.
- Baars, Bernard. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Baillargeon, Renée & Rose M. Scott & Zijing He. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, Vol. 14, No. 3, pp. 110–118.
- Bar-On, Dorit. (2009). First-person authority: dualism, constitutivism, and neo-expressivism. *Erkenntnis*, 71, pp. 53–71.
- Baron-Cohen, Simon. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.
- Bayne, Tim & Michelle Montague (Eds.). (2012). *Cognitive Phenomenology*. Oxford: Oxford University Press.
- Baz, Avner. (2010). Geach's 'refutation' of Austin revisited. *Canadian Journal of Philosophy*, Vol. 40, No. 1, pp. 41–62.
- Baz, Avner. (2012). *When Words Are Called For. A Defense of Ordinary Language Philosophy*. Cambridge, MA: Harvard University Press.
- Bermúdez, José L. (2003a). The domain of folk psychology. *Royal Institute of Philosophy Supplement*, 53, pp 25–48.
- Bermúdez, José Luis. (2003b). *Thinking without Words*. Oxford University Press.
- Bernecker, Sven. (1996). Externalism and the attitudinal component of self-knowledge. *Noûs*, Vol. 30, No. 2, pp. 262–275.
- Bohman, James. (2000). The importance of the second person: interpretation, practical knowledge and normative attitudes. In: Kögler, Hans Herbert & Karsten R. Stueber

- (eds). *Empathy & Agency. The Problem of Understanding in the Human Sciences*, pp. 222–242. Westview Press.
- Borge, Steffen. (2013). In defense of the received view. *Philosophical Psychology*. Vol. 26, No. 6, pp. 863–887.
- Brandom, Robert. (1994). *Making It Explicit. Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brandom, Robert. (2000). *Articulating Reasons. An Introduction to Inferentialism*. Harvard University Press.
- Breheny, Richard. (2006). Communication and folk psychology. *Mind & Language*, Vol. 21, No. 1, pp. 74–107.
- Bruner, Jerome. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.
- Byrne, Alex. (1998). Interpretivism. *European Review of Philosophy*, 3, pp. 199–223.
- Byrne, Alex. (2005). Introspection. *Philosophical Topics*, 33, pp. 79–104.
- Byrne, Alex. (2011). Knowing that I am thinking. In *Self-Knowledge*, Anthony Hatzimoyisis (ed.), pp. 105–124. Oxford: Oxford University Press.
- Byrne, Alex. (2012). Knowing what I want. *Consciousness and the Self: New Essays*. JeeLoo Liu & John Perry (eds.), pp. 165–183. Cambridge: Cambridge University Press.
- Byrne, Donn. (1971). *The Attraction Paradigm*. New York: Academic Press.
- Byrne, Richard W. & Andrew Whiten (Eds.). (1988). *Machiavellian Intelligence. Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Clarendon Press.
- Campbell-Meiklejohn, Daniel K. & Dominik R. Bach & Andreas Roepstorff & Raymond J. Dolan & Chris D. Firth. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, 20, pp. 1–6.
- Cappelen, Herman. (2012). *Philosophy without Intuitions*. Oxford: Oxford University Press.
- Carey, Susan. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.
- Carnap, Rudolf. (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- Carruthers, Peter. (1996). *Language, Thought and Consciousness. An Essay in Philosophical Psychology*. Cambridge: Cambridge University Press.
- Carruthers, Peter. (2009). How we know our own minds: the relationship between mindreading and metacognition. *Behavioural and Brain Sciences*, 32, pp. 121–182.
- Carruthers, Peter. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Cassam, Quassim. (2010). Judging, believing, and thinking. *Philosophical Issues*, 20, pp. 80–95.
- Cavell, Stanley. (1979). *The Claim of Reason. Wittgenstein, Skepticism, Morality, and Tragedy*. Oxford: Oxford University Press.
- Chalmers, David. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, pp. 200–219.
- Chartrand, Tanya L. & John A. Bargh. (1999). The chameleon effect: the perception-behavior link in social interaction. *Journal of Personality and Social Psychology*, Vol. 76, No.6, pp. 893–910.
- Cheng, Kai-Yuan. (2007). Semantic dispositionalism, idealization, and *ceteris paribus* clauses. *Minds & Machines*, 19, pp. 407–419.
- Child, William. (1994). *Causality, Interpretation, and the Mind*. Oxford: Clarendon Press.

- Christensen, David. (2009). Disagreement as evidence: the epistemology of controversy. *Philosophy Compass*, Vol. 4, No. 5, pp. 756–767.
- Chudnoff, Elijah. (2010). The nature of intuitive justification. *Philosophical Studies* 153, pp. 313–333.
- Churchland, Paul M. (1981). Eliminative materialism and propositional attitudes. *The Journal of Philosophy*, Vol. 78, No. 2, pp. 67–90.
- Clément, Fabrice. (2010). To trust or not to trust? Children’s social epistemology. *Review of Philosophy and Psychology*, Vol. 1, No. 4, pp. 531–549.
- Craig, Edward. (1990). *Knowledge and the State of Nature. An Essay in Conceptual Synthesis*. Clarendon Press, Oxford.
- Corbi, Josep E. (2009). First-person authority and self-knowledge as an achievement. *European Journal of Philosophy*, Vol. 18, No. 3, pp. 325–362.
- Cummins, Robert. (1996). *Representations, Targets, and Attitudes*. Cambridge, MA: MIT Press.
- Cummins, Robert. (2010). *The World in the Head*. Oxford: Oxford University Press.
- Currie, Gregory & Ian Ravenscroft. (2002). *Recreative minds: Imagination in philosophy and psychology*. Oxford: Oxford University Press.
- Cutter, Brian & Michael Tye. (2011). Tracking representationalism and the painfulness of pain. *Philosophical Issues*, Vol. 21, No. 1, pp. 90–109.
- Dahl, Robert. (1957). The Concept of Power. *Behavioral Science*, 2: 201–15.
- Davidson, Donald. (1963). Actions, reasons, and causes. *The Journal of Philosophy* 63, pp. 685–700.
- Davidson, Donald (1970). Mental events. In L. Foster and J. Swanson (Eds.), *Experience and Theory*. London: Duckworth. Reprinted in *Essays on Actions and Events*. Oxford: Oxford University Press, 1980, pp. 207–225.
- Davidson, Donald. (1973). Radical interpretation. *Dialectica*, 27, pp. 313–328.
- Davidson, Donald. (1982). Rational animals. *Dialectica*, 36, pp. 317–327.
- Davidson, Donald. (1990). Representation and interpretation. In Mohyeldin Said, Newton-Smith, Viale, and Wilkes, pp. 13–26.
- Dawkins, Richard. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- De Sousa, Ronald. (1998). Serendipity and desire. *Midwest Studies in Philosophy*, 22, pp. 120–134.
- Dennett, Daniel C. (1981). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press.
- Dennett, Daniel C. (1987). *The Intentional Stance*. Cambridge, MA: The MIT Press.
- Dennett, Daniel C. (1991a). *Consciousness Explained*. Little-Brown.
- Dennett, Daniel C. (1991b). Real patterns. *Journal of Philosophy*, 88, pp. 27–51.
- Dennett, Daniel C. (1996). Facing backwards on the problem of consciousness. *Journal of Consciousness Studies*, Vol. 3, No. 1, pp. 4-6.
- Descombes, Vincent. (2001). *The Mind’s Provisions. A Critique of Positivism*. Princeton University Press.
- Dretske, Fred. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Evans, Gareth. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Evans, Jonathan St. B. T. (2008). Dual-processing accounts of reasoning, judgement, and social cognition. *Annual Review of Psychology*, 59, pp. 255-278.
- Fisher, Justin C. (2006). *Pragmatic Conceptual Analysis*. Dissertation.  
[http://arizona.openrepository.com/arizona/bitstream/10150/195796/1/azu\\_etd\\_1968\\_sip1\\_m.pdf](http://arizona.openrepository.com/arizona/bitstream/10150/195796/1/azu_etd_1968_sip1_m.pdf). Downloaded at 18.09.2013
- Fodor, Jerry. (1975). *Language of Thought*. New York: Cromwell.

- Fodor, Jerry A. (1985). Fodor's guide to mental representation: the intelligent Auntie's vade-mecum. *Mind*, New Series, Vol. 94, No. 373, pp. 76–100.
- Fodor, Jerry. (1987). *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. A Bradford Book. The MIT Press.
- Frankfurt, Harry. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68, pp. 5–20.
- Frankfurt, Harry. (1992). The faintest passion. *Proceedings and Addressings of the American Philosophical Association*, Vol. 66, No. 3, pp. 5–16.
- Frankish, Keith. (2004). *Mind and Supermind*. Cambridge: Cambridge University Press.
- Friedrich, Daniel. (2012). The alluringness of desire. *Philosophical Explorations*, Vol. 15, No. 3, pp. 291–302.
- Fumerton, Richard. (2005). Speckled hens and the objects of acquaintance. *Philosophical Perspectives*, Vol. 19, No. 1, pp. 121–138.
- Furnham, Adrian. (1982). Personality and Activity Preference. *British Journal of Social and Clinical Psychology*, 20, pp. 57–68.
- Gallagher, Shaun. (2006). *How the Body Shapes the Mind*. Oxford: Clarendon Press.
- Garrels, Scott R. (2006). Imitation, mirror neurons, and mimetic desire: convergence between the mimetic theory of René Girard and empirical research on imitation. *Contagion: Journal of Violence, Mimesis, and Culture*, 12, pp. 47–86.
- Gauker, Christopher. (1994). *Thinking Out Loud. An Essay on the Relation between Thought and Language*. Princeton University Press.
- Gauker, Christopher. (1997). Domains of discourse. *Mind*, 106, pp. 1–32
- Gauker, Christopher. (2003a). *Words without Meaning*. Cambridge, Mass. MIT Press.
- Gauker, Christopher. (2003b). Attitudes without psychology. *Facta Philosophica*, 4, pp. 239–256.
- Gauker, Christopher. (2011). *Words and Images. An Essay on the Origin of Ideas*. Oxford: Oxford University Press.
- Gazzaniga, Michael S. (1985). *The Social Brain: Discovering the Networks of the Mind*. New York: Basic Books.
- Gendler, Tamar S. (2008). Alief and action (and reaction). *Mind & Language*, Vol. 23, No. 5, pp. 552–585.
- Gertler, Brie. (2009). Self-knowledge and the transparency of belief. In *Self-Knowledge*, Anthony Hatzimoysis (ed.). Oxford: Oxford University Press.
- Gibbs, Raymond Jr. (2001). Intentions as emergent products of social interactions. In B. F. Malle, L. J. Moses & D. A. Baldwin (eds.), *Intentions and Intentionality. Foundations of Social Cognition*, pp. 105–122. Cambridge, MA: MIT Press.
- Gilbert, Daniel T. & Douglas S. Krull & Patrick S. Malone. (1990). Unbelieving the unbelievable: some problems in the rejection of false information. *Journal of Personality and Social Psychology*, Vol. 59, No. 4, pp. 601–613.
- Gilbert, Daniel T. (1991). How mental systems believe. *American Psychologist*, Vol. 46, No. 2, pp. 107–119.
- Girard, René. (1976). *Deceit, Desire, and the Novel: Self and Other in Literary Structure*. Johns Hopkins University Press.
- Girard, René. (1995 [1977]). *Violence and the Sacred*. The Athlone press, London.
- Girard, René. (2004). Violence and religion: cause or effect? *The Hedgehog Review*, Vol. 6, No. 1, pp. 8–13.
- Glock, Hans-Johann. (2002). Does ontology exist? *Philosophy*, Vol. 77, No. 300, pp. 235–260.

- Godfrey-Smith, Peter. (2006). Mental Representation, Naturalism and Teleosemantics. In G. Macdonald and D. Papineau (eds.) *Teleosemantics. New Philosophical Essays*, pp. 42–68. Oxford: Oxford University Press.
- Goldman, Alvin. (1989). Interpretation psychologized. *Mind and Language*, 4, pp. 161–185.
- Goldman, Alvin. (2006). *Simulating Minds. The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
- Gopnik, Alison & Andrew N. Meltzoff. (1997). *Words, thoughts, and theories*. Cambridge: MIT Press.
- Gordon, Robert M. (1986). Folk psychology as simulation. *Mind and Language*, 1, pp. 158–171.
- Gordon, Robert M. (1996). Radical simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind*, pp. 11–21. Cambridge: Cambridge University Press.
- Gordon, Robert M. (2007). Ascent routines for propositional attitudes. *Synthese*, 159, pp. 151–165.
- Happé, Francesca. (1997). Central coherence and theory of mind in autism: reading homographs in context. *British Journal of Developmental Psychology*, 15, pp. 1–12.
- Haslanger, Sally. (1999). What knowledge is and what it ought to be: feminist values and normative epistemology. *Philosophical Perspectives*, 9, pp. 459–480.
- Haslanger, Sally. (2000). Gender and race. (What) are they? (What) we want them to be? *Noûs*, Vol. 34, No. 1, pp. 31–55.
- Haslanger, Sally. (2005). What are we talking about? The semantics and politics of social kinds. *Hypathia*, Vol. 20, No. 4, pp. 10–26.
- Haslanger, Sally. (2006). What good are our intuitions? *Aristotelian Society Supplementary Volume*, Vol. 80, No. 1, pp. 89–118.
- Hatfield, Elaine & John T. Cacioppo & Richard L. Rapson. (1993). *Emotional Contagion*. Cambridge: Cambridge University Press.
- Heal, Jane. (1998). Co-cognition and off-line simulation: two ways of understanding the simulation approach. *Mind & Language*, 13, pp. 477–498.
- Heal, Jane. (2002). On first-person authority. *Proceeding of the Aristotelian Society*, Vol. 102, No. 1, pp. 1–19.
- Heal, Jane. (2003). What are psychological concepts for? In *Mind, Reason and Imagination*, pp. 225–249. Cambridge: Cambridge University Press.
- Henrich, Joseph. (2009). Evolution of Costly Displays and Their Implication for Cultural Evolution. *Evolution and Human Behavior*, 30, pp. 244–260
- Henrich, Joseph & Francisco J. Gil-White. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22, pp. 165–196.
- Hirstein, William. (2005). *Brain-Fiction: Self-Deception and the Riddle of Confabulation*. Cambridge, MA: MIT Press.
- Hobbes, Thomas. (1996 [1641]). *Leviathan*. Oxford World's Classics. Oxford University Press.
- Horgan, Terry. (2012). From agentive phenomenology to cognitive phenomenology. In T. Bayne & M. Montague (Eds.), pp. 57–78. *Cognitive Phenomenology*. Oxford: Oxford University Press.
- Hunter, David A. (2011). Alienated belief. *dialectica*, Vol. 65, No. 2, pp. 221–240.
- Hutto, Daniel D. (1999). A cause for concern: reasons, causes, and explanations. *Philosophy and Phenomenological Research*, Vol. 59, No. 2, pp. 381–401.

- Hutto, Daniel D. (2004). The limits of spectatorial folk psychology. *Mind & Language*, Vol. 19, No. 5, pp. 548–573.
- Hutto, Daniel D. (2008). *Folk Psychological Narratives. The Sociocultural Basis of Understanding Reasons*. Cambridge: The MIT Press.
- Hutto, Daniel D. (2009). Lessons from Wittgenstein: elucidating folk psychology. *New Ideas in Philosophy*, 27, pp. 197–212.
- Hutto, Daniel D. (2011). Presumptuous naturalism: a cautionary tale. *American Philosophical Quarterly*, Vol. 48, No. 2, pp. 129–146.
- Hutto, Daniel D. (2013). Why believe in contentless beliefs? In N. Nottelman (Ed.), *New Essays on Belief: Structure, Constitution, and Content*, pp. 55–74. Palgrave MacMillan.
- Inzlicht, Michael & Talia Ben-Zeev. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11, pp. 365–371.
- Jackson, Frank & Philip Pettit. (1993). Folk belief and commonplace belief. *Mind & Language*, Vol. 8, No. 2, pp. 298–305.
- Jackson, Frank. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
- Johnston, Mark (2001). The authority of affect. *Philosophy and Phenomenological Research*, Vol. 63, No. 1, pp. 181–214.
- Kelman, Herbert C. (1958). Compliance, identification, and internalization: three processes of attitude change. *Journal of Conflict Resolution*, 2, pp. 51–60.
- Kenyon, Timothy. (2000). Indeterminacy and realism. In D. Ross, A. Brook & D. Thompson (Eds.), *Dennett's Philosophy. A Comprehensive Assessment*, pp. 77–94. Cambridge, MA: MIT Press.
- Kim, Jaegwon. (1998). *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation*. Cambridge: MIT Press.
- Klausen, Søren Harnow. (2008). The phenomenology of propositional attitudes. *Phenomenology and the Cognitive Sciences*, 7, pp. 445–462.
- Klein, Colin. (2007). An imperative theory of pain. *The Journal of Philosophy*, Vol. 104, No. 10, pp. 517–532.
- Knobe, Joshua. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, pp. 315–365.
- Kovács, Ágnes Melinda & Ernő Téglás & Ansgar Denis Endress. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, 330, pp. 1830–1834.
- Kripke, Saul A. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.
- La Caze, Marguerite. (2001). Envy and resentment. *Philosophical Explorations*, 1, pp. 31–45.
- Lahav, Ran. (1992). The amazing predictive power of folk psychology. *Australasian Journal of Philosophy*, 70, pp. 99–105.
- Laland, Kevin N. & Kim Sterelny. (2006). Seven reasons (not) to neglect niche construction. *Evolution*, Vol. 60, No. 9, pp. 1751–1762.
- Lance, Mark N. & O'Leary-Hawthorne, John. (1997). *The Grammar of Meaning. Normativity and Semantic Discourse*. Cambridge: Cambridge University Press.
- Lawlor, Krista. (2008). Knowing beliefs, seeking causes. *American Imago*, Vol. 65, No. 3, pp. 335–356.
- Lawlor, Krista. (2009). Knowing what one wants. *Philosophy and Phenomenological Research*, Vol. 79, No. 1, pp. 47–75.



- Levine, Joseph. (2012). On the phenomenology of thought. In T. Bayne & M. Montague (Eds.), *Cognitive Phenomenology*, pp. 103–120. Oxford: Oxford University Press.
- Lewis, David. (1966). An argument for the identity theory. *Journal of Philosophy* Vol. 63, No. 1, pp. 17–25.
- Lewis, David. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, Vol. 50, No. 3, pp. 249–258.
- Lukes, Steven. (2005). *Power: A Radical View*, 2nd expanded edition. London: Macmillan.
- Lycan, William G. (1988). *Judgement and Justification*. Cambridge: Cambridge University Press.
- Lycan, William G. (2008). Phenomenal intentionalities. *American Philosophical Quarterly*, 45, pp. 233–252.
- Lydon, John E. & David W. Jamieson & Mark P. Zanna. (1988). Interpersonal similarity and the social and the intellectual dimensions of first impressions. *Social Cognition*, 6, pp. 269–286.
- Machery, Edouard. (2009). *Doing without Concepts*. New York: Oxford University Press.
- MacKay, Gilbert & Adrienne Shaw. (2004). A comparative study of figurative language in children with autistic spectrum disorders. *Child Language, Teaching and Therapy*, 20, pp. 13–32.
- Malcolm, Norman. (1977). Thoughtless brutes. *Thought and Knowledge*. Ithaca: Cornell University Press, pp. 40–57.
- Malle, Bertram F. (2004). *How the Mind Explains Behaviour*. Cambridge, MA: MIT Press.
- Malle, Bertram F. & Joshua Knobe. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, pp. 101–121.
- Mameli, Matteo. (2001). Mindreading, mindshaping, and evolution. *Biology and Philosophy*, 16, pp. 597–628.
- Martinez, Manolo. (2011). Imperative content and the painfulness of pain, *Phenomenology and the Cognitive Sciences*, Vol. 10, No. 1, pp. 67–90.
- McGeer, Victoria. (2007). The regulative dimension of folk psychology. In *Folk Psychology Re-Assessed*, D.D. Hutto & M. Ratcliffe (Eds.), pp. 137–156. Dordrecht: Springer.
- McManus, Denis. (2000). Boghossian, Miller and Lewis on dispositional theories of meaning. *Mind & Language*, Vol. 15, No. 4, pp. 393–399.
- Medina, Jose. (2006). What's so special about self-knowledge? *Philosophical Studies*, 129, pp. 575–603.
- Meltzoff, Andrew. (1988). Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, Vol. 24, No. 4, pp. 470–476.
- Meltzoff, Andrew. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31: 838–850.
- Meltzoff, Andrew & Jean Decety. (2003). What imitation tells about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Phil. Trans. R. Soc. Lond. B*, 358, pp. 491–500.
- Meltzoff, Andrew & M. Keith Moore. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54, pp. 702–709.
- Melser, Derek. (2004). *The Act of Thinking*. Cambridge, MA: MIT Press.
- Melzer, Bernard N. & Gil Richard Musolf. (2002). Resentment and ressentiment. *Sociological Inquiry*, Vol. 72, No. 2, pp. 240–255.

- Millikan, Ruth G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Millikan, Ruth G. (2005). *Language: A Biological Model*. Oxford: Clarendon Press.
- Mohyeldin Said, K., Newton-Smith, W., Viale, R., and Wilkes, Kathleen. (Eds.). (1990). *Modelling the Mind*. Oxford: Oxford University Press.
- Mölder, Bruno. (2010). *Mind Ascribed: An Elaboration and Defense of Interpretivism*. John Benjamins.
- Montague, Michelle. (2007). Against propositionalism. *Noûs*, Vol. 41, No. 3, pp. 503–518.
- Moran, Richard. (1994). Interpretation theory and the first person. *The Philosophical Quarterly*, Vol. 44, No. 175, pp. 154–173.
- Moran, Richard. (2001). *Authority and Estrangement. An Essay on Self-Knowledge*. Princeton University Press.
- Moran, Richard. (2005). Getting told and being believed. *Philosophers' Imprint*, Vol. 5, No. 5, pp. 1–29.
- Morton, Adam. (1996). Folk psychology is not a predictive device. *Mind*, 105, pp. 119–137.
- Moss, Jessica. (2006). Pleasure and illusion in Plato. *Philosophy and Phenomenological Research*, Vol. 72, No. 3, pp. 503–535.
- Moyal-Sharrock, Danièle. (2009). Wittgenstein and the memory debate. *New Ideas in Psychology*, Vol. 27, No. 2, pp. 213–227.
- Nagel, Thomas. (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.
- Newcombe, Theodore M. (1961). *The Acquaintance Process*. New York: Holt, Rinehart and Winston.
- Nichols, Shaun. (2001). Mindreading and the cognitive architecture underlying altruistic motivation. *Mind & Language*, Vol. 16, No. 4, pp. 425–455.
- Nichols, Shaun & Stephen Stich. (2002). How to read your own mind: a cognitive theory of self-consciousness. In Q. Smith & A. Jokic (Eds.), *Consciousness: New Philosophical Essays*, pp. 157–200. Oxford: Oxford University Press.
- Nietzsche, Friedrich. (1888 [1887]). *Zur Genealogie der Moral*. Philipp Reclam jun. Stuttgart.
- O'Brien, Lucy. (2007) *Self-Knowing Agents*. Oxford: Oxford University Press.
- O'Leary-Hawthorne & Huw Price. (1996). How to stand up for non-cognitivists. *The Australasian Journal of Philosophy*, 74, pp. 275–292.
- Oddie, Graham. (2005). *Value, Reality, and Desire*. Oxford: Clarendon Press.
- Petty, Richard E. & Wegener, Duane T. (1998) Attitude change: Multiple roles for persuasion variables. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *The handbook of social psychology*, pp. 323–90. McGraw-Hill.
- Platts, Mark de Bretton. (1979). *Ways of Meaning*. London: Routledge & Kegan Paul.
- Price, Huw. (1988). *Facts and the Function of Truth*. Basil Blackwell.
- Price, Huw. (1989). Defending desire-as-belief. *Mind*, New Series, Vol. 98, No. 389, pp. 119–127.
- Price, Huw. (1992). Metaphysical pluralism. *The Journal of Philosophy*, Vol. 89, No. 8, pp. 387–409.
- Price, Huw. (2007). Quining naturalism. *The Journal of Philosophy*, Vol. 104, No. 8, pp. 375–402.
- Price, Huw. (2009). Metaphysics after Carnap: the ghost who walks? in David Chalmers et al. (Eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 320–346. Oxford University Press.

- Prinz, Jesse. (2012a). *Beyond Human Nature. How Culture and Experience Shape Our Lives*. Allen Lane.
- Prinz, Jesse. (2012b). *The Conscious Brain. How Attention Engenders Experience*. New York: Oxford University Press.
- Putnam, Hilary. (1992). *Renewing philosophy*. Cambridge, MA: Harvard University Press.
- Quine, Willard Van Orman. (1948). On what there is. *Review of Metaphysics*, Vol. 2, No. 5, pp 21–38.
- Quinn, Warren. (1995). Putting Rationality in Its Place. In R. Hursthouse, G. Lawrence & W. Quinn (Eds.), *Virtues and Reasons: Philippa Foot and Moral Theory*, pp. 189–190. Oxford: Clarendon.
- Rapp, Christof. (2010). Aristotle's Rhetoric. In *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/>, last accessed: February 01, 2012.
- Ratcliffe, Matthew. (2006). 'Folk psychology' is not folk psychology. *Phenomenology and the Cognitive Sciences*, 5, pp. 31–52.
- Ratcliffe, Matthew. (2007). *Rethinking Commonsense Psychology. A Critique of Folk Psychology, Theory of Mind and Simulation*. Palgrave MacMillan.
- Rochat, Philippe. (2009). *Others in Mind. Social Origins of Self-Consciousness*. Cambridge University Press.
- Rosenbaum, Milton E. (1986). The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51, pp. 1156–66.
- Rousseau, Jean-Jacques. (1984 [1754]). *A Discourse on Inequality*. New York: Penguin Books.
- Rudinow, Joel. (1978). Manipulation. *Ethics*, Vol. 88, No. 4, pp. 338–347.
- Rundblad, Gabriella & Dagmara Annaz. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, 14, pp. 29–46.
- Ryle, Gilbert. (1949). *The Concept of Mind*. Chicago: The University of Chicago Press.
- Ryle, Gilbert. (1962). Abstractions. *Dialogue*, Vol. 1, No. 1, pp. 5–16.
- Sally, David. (2000). A general theory of sympathy, mind-reading, and social interaction, with an application to Prisoner's Dilemma. *Social Science Information*, Vol. 39, No. 4, pp. 567–634.
- Scanlon, Thomas M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schapiro, Tamar. (2009). The nature of inclination. *Ethics*, 119, pp. 229–256.
- Schiffer, Stephen. (2003). *The Things We Mean*. Oxford: Clarendon Press.
- Scholl, Brian & Alan Leslie. (1999). Modularity, development, and 'Theory of Mind'. *Mind & Language*, Vol. 14, No. 1, pp. 131–153.
- Schroeder, Timothy. (2004). *Three Faces of Desire*. Oxford: Oxford University Press.
- Schueler, G. F. (1994). *Desire: Its Role in Practical Reasoning and the Explanation of Action*. Cambridge, MA: MIT Press.
- Schwitzgebel, Eric. (2001). In-between believing. *The Philosophical Quarterly*, Vol. 51, No. 202, pp. 76–82.
- Schwitzgebel, Eric. (2002). A phenomenal, dispositional account of belief. *Noûs*, Vol. 36, No. 2, pp. 249–75.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or, the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91, pp. 531–53.

- Schwitzgebel, Eric. (2013). A dispositional approach to attitudes: thinking outside of the belief box. In N. Nottelman (Ed.), *New Essays on Belief: Structure, Constitution, and Content*, pp. 75–99. Palgrave MacMillan.
- Searle, John R. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.
- Segal, Gabriel. (2000). *A Slim Book About Narrow Content*. Cambridge, MA: MIT Press.
- Sellars, Wilfrid. (1974). Meaning as functional classification. *Synthese*, 27, pp. 417–437.
- Sellars, Wilfrid. (1997). *Empiricism and the Philosophy of Mind*. Cambridge: Harvard University Press.
- Senju, A., V. Southgate, Y. Miura, T. Matsui, T. Hasegawa, Y. Tojo, H. Osanai and G. Csibra (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*, 22, pp. 353–360.
- Shields, Christopher. (2012). On behalf of cognitive qualia. In T. Bayne & M. Montague (Eds.), *Cognitive Phenomenology*, pp. 215–235. Oxford: Oxford University Press.
- Silins, Nicholas. (2012). Judgement as a guide to belief. In D. Smithies & D. Stoljar (Eds.), *Introspection and Consciousness*, pp. Oxford: Oxford University Press.
- Skyrms, Brian. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
- Smith, Adam. (1984). *The Theory of Moral Sentiments*. Liberty Fund, Indianapolis.
- Smith, Michael. (1987). The Humean theory of motivation. *Mind*, New Series, Vol. 96, No. 381, pp. 36–61.
- Smithies, Declan. (2012). A simple theory of introspection. In D. Smithies & D. Stoljar (Eds.), *Introspection and Consciousness*, pp. 259–294. Oxford: Oxford University Press.
- Spencer, Steven J. & Claude M. Steele & Diane M. Quinn. (1999). Stereotype threat and women’s math performance. *Journal of Experimental Social Psychology*, 35, pp. 4–28.
- Sperber, Dan. (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.
- Sperber, Dan. (2000). An objection to the memetic approach to culture. In R. Aunger (Ed.), *Darwinizing Culture: The Status of Memetics as a Science*, pp. 163–173. Oxford University Press.
- Sperber, Dan & Fabrice Clement & Christophe Heintz & Olivier Mascaro & Hugo Mercier & Gloria Origgi & Deirdre Wilson. (2010). Epistemic Vigilance. *Mind & Language*, Vol. 25, No. 4, pp. 359–393.
- Stalnaker, Robert C. (1987). *Inquiry*. Cambridge, MA: MIT Press.
- Stalnaker, Robert C. (1998). On the representation of context. *Journal of Logic, Language and Information*, 7, pp. 3–19.
- Sterelny, Kim. (2003). *Thought in a Hostile World. The Evolution of Human Cognition*. Blackwell Publishing.
- Sterelny, Kim. (2007). Social intelligence, human intelligence and niche construction. *Philosophical Transactions of the Royal Society B*, 362, pp. 719–730.
- Steward, Helen. (1997). *Ontology of Mind. Events, Processes, and States*. Oxford: Oxford University Press.
- Stich, Stephen. (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- Stich, Stephen. (1996). *Deconstructing the Mind*. New York: Oxford University Press.

- Stich, Stephen & Shaun Nichols. (1993). Folk psychology: simulation or tacit theory. *Mind and Language*, 7, pp. 35–71.
- Stich, Stephen P. & Ravenscroft Ian. (1996). What is folk psychology? In S. P. Stich, *Deconstructing the Mind*, pp. 115–135. Oxford: Oxford University Press.
- Stich, Stephen & Shaun Nichols. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.
- Stone, Tony & Martin Davies. (1996). The mental simulation debate: a progress report. In P. Carruthers & P.K. Smith (Eds.), *Theories of theories of mind*, pp. 119–137. Cambridge: Cambridge University Press.
- Stout, Rowland. (2006). *The Inner Life of a Rational Agent. In Defence of Philosophical Behaviourism*. Edinburgh: Edinburgh University Press.
- Strawson, Galen. (1994). *Mental Reality*. Cambridge, MA: MIT Press.
- Strawson, Peter F. (1974). *Freedom and Resentment and Other Essays*. London: Methuen.
- Tager-Flusberg, Helen. (1994). Dissociations in form and function in the acquisition of language by autistic children. In H. Tager-Flusberg (Ed.), *Constraints on Language Acquisition: Studies of Atypical Children*, Hillsdale, NJ: Erlbaum, pp. 175–194.
- Tanney, Julia. (1995). Why reasons may not be causes. *Mind & Language*, Vol. 10, No. 1/2, pp. 103–126.
- Tanney, Julia. (2005). Reason-explanation and the contents of the mind. *Ratio (new series)*, 18, pp. 338–351.
- Tanney, Julia. (2009). Reasons as non-causal, context-placing explanations. In Constantine Sandis (Ed.), *New Essays on the Explanation of Action*, pp. 94–111. Palgrave MacMillan.
- Träuble, Birgit & Vesna Marinović & Sabina Pauen. (2010). Early theory of mind competencies: do infants understand others' beliefs. *Infancy*, Vol. 15, No. 4, pp. 434–444.
- Van Cleave, Matthew & Christopher Gauker. (2010). Linguistic Practice and False-Belief Tasks. – *Mind & Language*, Vol. 25, No. 3, pp. 298–328.
- Van Fraassen, Bas. (1981). *The Scientific Image*. Oxford: Oxford University Press.
- Velleman, J. David. (1992). What happens when someone acts? *Mind*, 101, pp. 461–481.
- Voerman, Sander. (2011). On the disconfirmation of practical judgements. *Logique et Analyse*, Vol. 54, No. 216.
- Wellman, Henry. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- Wicker, Bruno & Christian Keysers & Jane Plailly & Jean-Pierre Royet & Vittorio Gallese & Giacomo Rizzolatti. (2003). Both of us disgusted in my *insula*: the common neural basis of seeing and feeling disgust. *Neuron*, 40, pp. 655–664.
- Wikerson, William S. (2001). Simulation, theory, and the frame problem: the interpretive movement. *Philosophical Psychology*, 14, pp. 141–153.
- Williams, Bernard. (2002). *Truth and Truthfulness. An Essay in Genealogy*. Princeton University Press.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy. (2006). Conceptual truth. *The Aristotelian Society Supplementary Volume*, 80, pp. 1–41.
- Willis, Janine & Alexander Todorov. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, Vol. 17, No. 7, pp. 592–598.

- Wilson, Robert A. (2003). Intentionality and phenomenology. *Pacific Philosophical Quarterly*, 84, pp. 413–431.
- Wilson, Timothy D. & Suzanne LaFleur. (1995). Knowing what you'll do: effects of analyzing reasons on self-prediction. *Journal of Personality and Social Psychology*, 68, pp. 21–35.
- Wilson, Timothy D. & Daniel T. Gilbert. (2005). Affective forecasting. Knowing what to want. *Current Directions in Psychological Science*, Vol. 14, No. 3, pp. 131–134.
- Wittgenstein, Ludwig. (1980). *Remarks on the Philosophy of Psychology, Vol. I*. Oxford: Blackwell.
- Wittgenstein, Ludwig. (1984). *Tractatus Logico-Philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Suhrkamp Taschenbuch, Wissenschaft 501.
- Wohlschläger, Andreas & Harold Bekkering. (2002). The role of objects in imitation. In Stamenov, Maksim I. & Vittorio Gallese (Eds.), *Mirror Neurons and the Evolution of Brain and Language*, pp. 101–114. John Benjamins Publishing Company.
- Woodward, James. (2009). Why do people cooperate as much as they do? In C. Mantzavinos (Ed.), *Philosophy of the Social Sciences: Philosophical Theory and the Scientific Practice*, pp. 219–265. Cambridge: Cambridge University Press.
- Zawidzki, Tadeusz. (2008). The function of folk psychology: mind reading or mind shaping? *Philosophical Explorations*, 11, pp. 193–210.
- Zawidzki, Tadeusz. (2013). *Mindshaping: Linchpin of the human socio-cognitive syndrome*. Cambridge, MA: MIT Press.
- Zimmerman, Aaron. (2007). The nature of belief. *Journal of Consciousness Studies*, Vol. 14, No. 11, pp. 62–82.

## SUMMARY IN ESTONIAN

### Uskumuste ja soovide kommunikatiivne tähtsus

Uskumuste ja soovide omistamise võime – rahvapsühholoogia – on viimastel kümnenditel olnud nii psühholoogias kui filosoofias intensiivse teoreetilise huvi objektiks. Peaks olema selge, et teadmine iseenda ja teiste hoiakute kohta on inimese jaoks tähtis ning et ilma sellise teadmiseta inimest sotsiaalse olendina ilmselt ei eksisteeriks. Selgitamist vajab aga küsimus, milles see tähtsus täpsemalt seisneb. Uskumuste ja soovide omistamise tähtsus on mõistetav selle peamiste kasulike tagajärgede kaudu omistaja jaoks. Siinse väitekirja esmaseks taotluseks on näidata, et nende tagajärgede hulka kuuluvad teatavad viisid, kuidas omistaja enda omistustest mõjutatud on, millest on senini rahvapsühholoogia uurimises üle vaadatud.

Väitekiri ei piirdu aga rahvapsühholoogiliste omistuste tagajärgede kirjeldamisega. Viimaseid arvesse võtnuna saab juba küsida, mis hoiakute omistamise juures teeb just sellised tagajärjed võimalikuks. Nõnda jõuame mu töö teise eesmärgini, milleks on uskumuste ja soovide omistustingimuste sõnastamine viisil, mis aitaks mõista, miks on omistustel just sellised tagajärjed, nagu neil on. Väidan, et rahvapsühholoogia peamiste kasulike tagajärgede seletamiseks tuleb meil konstrueerida omistustingimused nii, et omistused oleksid mõistetud teatavat laadi kõneaktide kaudu ja et nad tugineksid tõsiasjadele isiku käitumise ja heaolu kohta.

Uurimaks seda, mis on uskumuste ja soovide omistamise tagajärjed, tuginen tavakeele filosoofiast (Ludwig Wittgenstein, Gilbert Ryle) inspireeritud lähenemisele, mis vaatleb mõiste – antud juhul siis uskumuse ja soovi mõistete – kasutuse tagajärge igapäevases keelepraktikas. Neid tagajärge arvessevõtivate omistustingimuste konstrueerimisel otsin aga abi pragmaatiliselt mõisteanalüüsilt (Justin Fisher), mis võimaldab näha, millised peaksid olema mõiste rakendamise tingimused, et rakendusest tulenevad tagajärjed võimalikud oleksid. Tavakeele filosoofiast inspireeritud lähenemine aitab seega selgitada uskumuste ja soovide omistamise kõige olulisemaid tagajärge ning pragmaatiline kontseptuaalne analüüs laseb näha nende tagajärgede võimalikkuse tingimusi kõige optimaalsemal kombel. Väitekirja probleemipüstituse ning metodoloogilise inspiratsiooni avan ma esimeses, sissejuhatavas peatükis.

Teises peatükis uurin ma olemasolevaid käsitlusi rahvapsühholoogiliste omistuste peamiste tagajärgede kohta. Seal sõnastan ma ka oma keskse teesi. Selle kohaselt on sotsiaalse tunnetuse uurimises, kus teooriate rõhuasetus on olnud rahvapsühholoogia seletaval, ennustaval ning vähemal määral ka selle ettekirjutaval rollil, üks oluline aspekt uskumuste ja soovide tähtsuses senini tähelepanuta jäänud. Selleks aspektiks on need eriomased viisid, kuidas uskumuste ja soovide omistused omistajat mõjutavad (ma nimetan seda töö vältel ka rahvapsühholoogilistest omistustest tulenevaks “mõjutatavuseks”). Uskumuse omistamise korral on võimalik omistajal sellega nõustuda või mitte ning soovi omistamine annab võimaluse seda soovi kas heaks kiita või halvaks

panna. See tähendab, et soovide ja uskumuste omistamine ei seleta ega ennusta kõigest isikute tegusid ja teisi vaimuseisundeid, vaid annab omistajale ka võimaluse omistatud hoiakutele vastata ja nende tõesust (uskumused) või sobivust (soovid) hinnata.

Nende esmaste reaktsioonide peale ehituvad edasised võimalikud käigud sotsiaalses interaktsioonis. Nende käikude analüüs on kolmanda peatüki ülesandeks. Ma keskendun seal ainult mõningatele näidetele, püüdmata ammendavalt kirjeldada kõiki võimalusi, mis esmastest reaktsioonidest tulenevad. Eesmärk on näidata, et rahvapsühholoogiast tulenev mõjutatavus on süstemaatiliselt artikleeritav ning selle alguspunktiks on vahetud reaktsioonid omistatud uskumustele ja soovidele. Vaatluse all on kolmandas peatükis ennekõike need edasised käigud, mis sõltuvad hoiaku omistaja ja selle, kellele omistatakse, sotsiaalsest vahekorra, ning olukorrad, kus omistaja ei nõustu teise uskumusega või ei kiida heaks teise soovi. Juhul kui esimene ei nõustu teise hoiakuga, teine on aga esimesest sotsiaalselt mõjuvõimsamal positsioonil, siis on ta kallutatud esimese hoiakut üle võtma. Kuna aga hoiaku ülevõtmine on tingitud ebavõrdsest vahekorra, siis saadab seda loomuliku emotsionaalse reaktsioonina vimm teise suhtes. Kui aga omistaja pole teise suhtes, kelle hoiakuga ta ei nõustu, nõrgemal positsioonil, on suurem võimalus, et ta püüab teise hoiakut kas manipulatsiooni või vaidluse kaudu muuta. Omistatud hoiak, millega omistaja ei nõustu, mõjub kõigil neil juhtudel väljakutsena omistaja subjektiivsele perspektiivile, samas kui kooskõlaline hoiak kinnitab seda, ning omistaja edasised reaktsioonid sõltuvad tema ja teise isiku vahekorra.

Kui teine ja kolmas peatükk tegelevad rahvapsühholoogia kasulike tagajärgede kirjeldamisega, siis neljandas peatükis pöördun väitekirja teise eesmärgi poole. Taotluseks on sõnastada uskumuste ja soovide sellised omistustingimused, mis seletaksid kõige optimaalsemalt, kuidas mainitud rahvapsühholoogiale olemuslikud reaktsioonid (nõustumine või mittenõustumine, heakskiit või halvaksapanu) võimalikud on. Rahuldav omistustingimuste kirjeldus peaks lisaks neile reaktsioonidele seletama ka omistuste seletavat, ennustavat ja ettekirjutavat rolli. Väidan, et hoiakute omistustingimuste konstrueerimiseks tasub alustada tõsiasjast, et uskumustele ja soovidele olemuslikud reaktsioonid peegeldavad neid reaktsioone, mis on kohased ka väidetele ja käsklustele. Käsku määratlen ma antud juhul natuke laiemalt, kui seda tavakasutuses mõistetakse, nimelt kõneaktina, mis hindab mingit võimalikku asjadeseisu täideviimisväärsena. Nii uskumuste kui ka väidete puhul on kohaseks reaktsiooniks nendega kas nõustuda või mitte ning nii soove kui ka käske on võimalik heaks kiita või halvaks panna. Seetõttu tasub hoiakute omistustingimusi mõista viitega vastavatele kõneaktidele. Olemasolevatest autoritest leiab selle võimaluse vaid Christopher Gaukerilt. Tema vaimuseisundite kommunikatiivse kontseptsiooni kohaselt on uskumuse omistamine kellelegi sama mis väitmine selle isiku eest. Soovi omistamine on aga samastatav kellegi eest käsu andmisega. Kuna Gaukeri teooria seletab ainsana rahuldavalt uskumuste ja soovide suhtes kohaseid reaktsioone, siis kaitsen seda kui omistustingimuste esialgset kirjeldust, mis annab meile hoiakuomistuse tarvilikud tingimused.



Viies peatükk võtab aga kriitilise hoiaku kommunikatiivse kontseptsiooni suhtes. Argumenteerin seal, et Gaukeri teooria seisab ühe tõsise probleemi ees. Kuigi ta annab meile esialgse omistustingimuste kirjelduse, ei suuda ta rahuldavalt seletada, miks üldse uskumuse ja soovi mõisteid meile vaja on. Teise eest tehtavate väidete ja käskluste kasutus keeles paistab juba neidsamu tulemusi andvat, mis uskumustelt ja soovidelt eeldatud on. Kommunikatiivne kontseptsioon vajab seega täiendamist, et oleks seletatud, miks teatud olukordades enam teise isiku eest tehtud kõneaktidest ei piisa, vaid ka miks on vaja just rääkida spetsiifiliselt uskumustest ja soovidest. See tähendab ühtlasi, et Gaukeri algne teooria ei võimalda meil mõista rahvapsühholoogia seletavat rolli. Väitekirjas vastan sellele väljakutsele nii, et soovide puhul seletan nende postuleerimise tarvilikkust olukordadega, kus küsimuse all on, kas teise isiku eest antud käsu täitmine sellele isikule ka subjektiivset rahuldust pakub. Rahuldust on antud juhul mõistetud lihtsalt positiivse valentsiga tundena mingi asjadeseisu suhtes. Sooviomistus ühendab endas seega käsu andmist teise eest ning oletust, et selle täideviimine teist rahuldab (kalduvust tunda rahuldust mingi asjadeseisu suhtes nimetan ma väitekirjas “hedooniliseks seadumuseks”), isegi kui teine seda käsklust ise kuuldavale ei too. Uskumustest rääkimise tarvilikkus ilmneb aga olukordades, kus on vaja midagi isiku eest väita, arvestades tema käitumist, kuigi ta ise seda ei ütle. Sellistes olukordades on meil võimalik kohelda isikut, justkui ta oleks midagi väitnud, et siis väljendada meiepoolset nõustumist või mittenõustumist temaga. Uskumuseomistus võimaldab meil laiendada väidetele kohaseid reaktsioone ka mitteverbaalsele käitumisele. Soovide ja uskumuste omistamine saab seega vajalikuks siis, kui omistajal on vaja teise isiku eest midagi väita või käskida, võttes arvesse teatavaid hedoonilisi ja käitumuslikke tõsiasju tolle isiku kohta, ka siis, kui too isik vastavat kõneakti ise ei soorita. Uskumuste ja soovide omistamine laiendab keelelist kommunikatsiooni eriomasel viisil, koheldes isikuid, kellele neid hoiakuid omistatakse, virtuaalsete vestluspartneritena, justkui nad oleksid midagi väitnud või käskinud. Käitumuslikke ja hedoonilisi tõsiasju arvessevõttev kommunikatiivne kontseptsioon annab seega meile omistustingimuste rahuldava analüüsi.

Väitekirja kuues ja seitsmes peatükk vastavad kahele võimalikule kahtlusele täiendatud kommunikatiivse kontseptsiooni suhtes. Esimene neist, millele vastan kuuendas peatükis, puudutab vaimuseisundite ontoloogiat. Pakutud versioon kommunikatiivsest kontseptsioonist on esmajoones teooria selle kohta, mis kuulub konstitutiivselt uskumuste ja soovide omistuste juurde, ning ei paista omavat vahetuid järelemeid nende hoiakute endi loomuse kohta. Sellegipoolest saab teatava eelduse omaks võtu korral argumenteerida, et modifitseeritud kommunikatiivne kontseptsioon on ontoloogiliselt kohustunud tunnistama kõigest käitumuslike ja hedooniliste seadumuste olemasolu ning ei midagi enam. Kuni inimestel on säärased seadumused – ning on mõistlik eeldada, et neil need on – on neil ka uskumused ja soovid. Omaks võetavaks eelduseks on, et rahvapsühholoogia ontoloogiliste kohustuste hulka on tarvis hõlmata vaid neid eeldusi uskumuste ja soovide kohta, milleta nende omistamine enam oma tavapärasest rolli mängida ei saaks. Kuna seadumuste eeldamisest piisab juba

selleks, et rahvapsühholoogilised omistused oma tagajärgi (mõjutatavus, seletused, ennustused, ettekirjutused) omaksid, siis pole vaja rahvapsühholoogia ontoloogiliste kohustuste hulka rohkemat lugeda. Saab seega väita, et kuigi hoiakute omistustel on võrdlemisi keerukas struktuur, hõlmates teise eest tehtavat kõneakti, ei kohusta selle tunnistamine meid väitma midagi ontoloogiliselt ekstravagantset (ekstravagantne oleks näiteks seisukoht, mille kohaselt hoiaku olemasolu on mingil moel omistajast sõltuv). Rahvapsühholoogia on seetõttu võrdlemisi immunne teaduslikust psühholoogiast tulenevate empiiriliste avastuste suhtes (ja vastupidi).

Teine võimalik kahtlus tuleneb sellest, et kommunikatiivne kontseptsioon kohtleb hoiakute omistuse paradigmana olukordi, mil üks isik omistab *teisele* uskumusi ja soove. Rahuldav rahvapsühholoogia käsitlus peab aga rakenduma ka eneseleomistustele. Viimaste puhul tundub isiku suhe omistatud hoiakutesse olevat mingil moel vahetum kui suhe kolmanda isiku perspektiivilt. Ühtlasi tunduvad eneseleomistused olevat episteemiliselt privilegeeritud: isik paistab olevat teistest paremal positsioonil selle teadmiseks, mis hoiakud tal on. Väitekirja seitsmendas peatükis argumenteerin, et kommunikatiivne kontseptsioon suudab seletada esimest intuitsiooni. Ta ei pea aga seletama teist, kuna hoiakute eneseleomistused pole tegelikult episteemiliselt privilegeeritud. Eneseleomistuste kogemuslikku eripära seletan seeläbi, et erinevalt omistustest teistele isikutele ei kujuta need väitmist või käskimist *kellegi eest*, vaid tegu on lihtsalt väite või käsklusega. Kui keegi siiralt väidab või käsib midagi, kas avalikult või vaikimisi, kvalifitseerub see juba vastavalt uskumuse või soovi eneseleomistuseks. Sellest tuleneb ka intuitsioon, et isiku uskumused ja soovid on talle endale mingil kombel vahetumalt antud kui teistele. Erinevalt omistustest teistele, on eneseleomistuste eripäraks ka see, et kohased reaktsioonid nendele on piiratumad. Nii oma uskumusega mittenõustumine kui ka halvakspanu oma soovi suhtes märgivad isiku võõrandumist oma vaimuelust. Esimese ja kolmanda isiku perspektiivi sellised erinevused ei implitseeri aga seda, et eneseleomistused oleksid episteemiliselt privilegeeritud. Oma uskumuste ja soovide teadmiseks on isikul vaja teada oma käitumuslikke ja hedoonilisi seadumusi ning viimastele pole tal tagatud teistest paremat ligipääsu. Kui mingist privileegist rääkida saab, siis seisneb see vaid võimalikus tõsiasjas, et isikul on üldiselt rohkem informatsiooni oma seadumuste kohta.

Väitekirja järeldused motiveerivad ka edasisi uurimisküsimusi. Esiteks, kuna ma võtsin omaks kommunikatiivsest kontseptsioonist tuleneva arusaama, et teisele isikule hoiaku omistamine tähendab tema eest kõnelemist, siis sellest tulenevalt väärib edasist uurimist rahvapsühholoogia eetilise mõde. Juhul kui omistaja ja see, kellele hoiakut omistatakse, on näiteks omavahel sotsiaalselt ebavõrdses suhtes, võib see mõjutada omistuse sobivust, kuna sellistes olukordades ei pruugi teise eest kõnelemine olla süütute tagajärgedega. Teine lahtine uurimisküsimus tuleneb sellest, et sõnad “uskumus” ja “soov” võivad olla mitmetähenduslikud. Oma väitekirjas avasin ma ühe võimaliku tähenduse, aga jätsin lahtiseks võimaluse kontseptualiseerida uskumusi ja soove teistel

viisidel, mis samuti kasulikuks võivad osutada. Selle tunnistamine laseb ühtlasi uurida kultuurilisi ja ajaloolisi erinevusi hoiakute mõistmises, millest analüütilises vaimufilosoofias on võrdlemisi üle vaadatud. Kolmandaks tasub välja tuua rahvapsühholoogilistele omistustele kohaste reaktsioonide edasiuurimist, nii seda, mis puudutab nende struktuuri kui ka nende evolutsioonilist ajalugu. Eriti just reaktsioonide evolutsioonilise ajaloo avamine on vajalik selleks, et täpsemalt aru saada, kuidas meie võime hoiakuid hinnata on välja kujunenud. Nõnda on võimalik suhestada väitekirjas öeldut ka rahvapsühholoogia empiiriliste käsitlustega. Neljandaks vajavad edasist uurimist pragmaatilise kontseptuaalse analüüsi metodoloogilised alused, mille näol on tegemist võrdlemisi ebatüüpilise lähenemisega praeguses filosoofias, kuid mille potentsiaal ulatub kaugemale uskumuse ja soovi mõiste analüüsist. Lahtised küsimused on näiteks, kui peenekoeliselt tuleks eristada mõistekasutuse tagajärgi ning mis on sobivate mõiste omistustingimuste hindamise juures analüüsija intuitsioonide ja empiirilise tõendusmaterjali vahekord. Viiendaks, eneseleomistuste käsitlus jättis lahtiseks, et meil võib olla privilegeeritud ligipääs oma väidete ja käskluste sisule, mis on uskumuste ja soovide eneseleomistuste osaks. Kuna tegemist on kõneaktidega, mis erinevalt uskumustest ja soovidest on episoodilised, siis võimalus, et isikutel on neile privilegeeritud ligipääs, tundub intuiitiivselt võrdlemisi tõenäoline. Kas sellest järeldub midagi ka uskumuste ja soovide eneseleomistuste epistemoloogia jaoks, on lahtine küsimus. Lahtiseks jäi ka kujutlusvõime roll eneseleomistuste epistemoloogias, mille puhul võiks eeldada, et oskus kujutleda oma käitumist ja emotsionaalseid reaktsioone seoses omistatava hoiaku sisuga võib pakkuda teatava erilise viisi oma hoiakute teadasaamiseks. Samas on kaheldav, et pelk kujutlemine tagaks hoiaku olemasolu ja sellest johtuva tugeva episteemilise privileegi.

## CURRICULUM VITAE

**Name:** Uku Tooming  
**Date of Birth:** 21.01.1986  
**Contact:** Tiigi 19–11, 50410 Tartu, Estonia  
uku.tooming@gmail.com

### Education:

2010–2014 University of Tartu, PhD studies in philosophy  
Thesis title: “The Communicative Significance of Beliefs and Desires”, Bruno Mölder as supervisor  
Visiting researcher in the Radboud University Nijmegen,  
February – June, 2013  
Visiting researcher in the University of Hertfordshire, October –  
December, 2012

2007–2010 University of Tartu, MA degree *cum laude*, anthropology of  
religion  
Thesis title: “Religion’s Folk Psychological Context”,  
Jaan Kivistik as supervisor  
Exchange student in the University of Helsinki, Faculty of  
Theology, 2008/2009

2004–2007 University of Tartu, Bachelor’s Degree *cum laude*, philosophy  
Thesis title: “Explanation and Prediction in Folk Psychology”,  
Bruno Mölder as supervisor

### Research Interests:

Folk psychology, intentionality, self-knowledge, desires, anthropology of religion.

### Courses Taught:

2013 *Enesekohane teadmine*, University of Tartu  
2011 *Sissejuhatus vaimufilosoofiasse* (taught together with Taavi Laanpere), University of Tartu

### Publications:

Forthcoming “Pleasures of the Communicative Conception”. *Grazer Philosophische Studien*.  
“Without pretense: a critique of Goldman’s model of simulation”. *Phenomenology and the Cognitive Sciences*.

2013 “Review of Julia Tanney’s “Rules, Reason, and Self-knowledge”. *Philosophical Inquiries*, 1, pp. 1–5.

2012 “Enesetõlgendus ja soovid”. [Self-interpretation and desires.] *Akadeemia* 3, pp. 1099 – 1127.  
“Religion’s folk psychological context”. In: Jackelén, Antje; Smedes, Taede; Fuller, Michael; Evers, Dirk (Eds.) *Studies in*

*Science and Theology, No. 13: European Society For the Study of Science and Theology Conference, University of Edinburgh, 07.04–10.04, 2010.*

- 2011 “Intuitsioonidest eksperimentaalses filosoofias” [About intuitions in experimental philosophy]. In: Mölder, Bruno; Kangilaski, Jaan (Eds.). *Filosoofia ja analüüs: analüütilise filosoofia seminar 20*. Tartu: EYS Veljesto Kirjastus
- 2010 “Kuidas saab jumal tahta, teha ja tunda?” [How can God want, do or feel anything?] *Akadeemia* 6, pp. 447–472.

**Presentations:**

- 2013 “Läbipaistvuse läbikukkumine” [Failure of Transparency], IX Annual Estonian Philosophy Conference, December 14.  
“Second-personal attributions of mental states”, The Second-Person Perspective in Science and Humanities, Oxford, July 19.  
“What to do with beliefs and desires”, Mind & Action Seminar, Radboud University Nijmegen, March 15.
- 2012 “The communicative role of folk psychology”, Postgrad Seminar, University of Hertfordshire, November 26.  
“When do we need to talk about desires” [Poster], GAP.8, Konstanz, September 17–20.  
“First-person mindshaping” [Poster], Social Cognition, Engagement, and the Second-Person-Perspective, Cologne, May 25–27.
- 2011 “The role of enactive imagination in Goldman’s simulation theory” [Poster], Understanding Other Minds – Embodied Interaction and Higher-Order Reasoning, Bochum, September 20.  
“About the function of folk psychology”, Seventh European Congress of Analytic Philosophy, September 1–6,
- 2010 “Religion’s folk psychological context”, European Conference on Science and Theology, Edinburgh, April 7–11.

**Awards:**

- 2013 Akadeemia Silver Award for the article “Enesetõlgendus ja soovid”. [Self-interpretation and desires.] *Akadeemia* 3, pp. 1099–1127.
- 2011 Akadeemia Silver Award for the article “Kuidas saab jumal tahta, teha ja tunda?” [How can God want, do or feel anything?] *Akadeemia* 6, pp. 447–472.

## ELULOOKIRJELDUS

**Nimi:** Uku Tooming  
**Sünniaeg:** 21.01.1986  
**Kontakt:** Tiigi 19–11, 50410 Tartu, Eesti  
uku.tooming@gmail.com

### Hariduskäik:

- 2010–2014 Tartu Ülikool, doktoriõpingud filosoofias  
“The Communicative Significance of Beliefs and Desires”,  
juhendaja Bruno Mölder  
Külalisuuriija Nijmegeni Radboud Ülikoolis, veebruar–juuni,  
2013  
Külalisuuriija Hertfordshire’i Ülikoolis, oktoober–detsember,  
2012
- 2007–2010 Tartu Ülikool, MA religiooniantropoloogias (*cum laude*)  
Magistritöö: “Religion’s Folk Psychological Context”,  
juhendaja Jaan Kivistik  
Erasmuse vahetusüliõpilane Helsingi Ülikoolis, usuteaduskond,  
2008/2009 õppeaastal.
- 2004–2007 Tartu Ülikool, BA filosoofias (*cum laude*)  
Thesis title: “Explanation and Prediction in Folk Psychology”,  
juhendaja Bruno Mölder

### Peamised uurimisvaldkonnad:

Rahvapühholoogia, intentsionaalsus, enesekohane teadmine, soovid,  
religiooniantropoloogia.

### Õpetatud kursusi:

- 2013 *Enesekohane teadmine*, Tartu Ülikool  
2011 *Sissejuhatus vaimufilosoofiasse* (õpetatud kahasse Taavi Laan-  
perega), Tartu Ülikool

### Publikatsioonid:

- Ilmumas “Pleasures of the Communicative Conception”. *Grazer Philosophische Studien*.  
“Without pretense: a critique of Goldman’s model of simulation”. *Phenomenology and the Cognitive Sciences*.
- 2013 “Review of Julia Tanney’s “Rules, Reason, and Self-knowledge.” *Philosophical Inquiries*, 1, 1–5.
- 2012 “Enesetõlgendus ja soovid”. *Akadeemia*, 3, 1099–1127.  
“Religion’s folk psychological context”. Kogumikus: Jackelén, Antje; Smedes, Taede; Fuller, Michael; Evers, Dirk (Eds.) *Studies in Science and Theology, No. 13: European Society For*

*the Study of Science and Theology Conference, University of Edinburgh, 07.04–10.04, 2010.*

- 2011 “Intuitsioonidest eksperimentaalses filosoofias”. Kogumikus: Mölder, Bruno; Kangilaski, Jaan (Eds.). *Filosoofia ja analüüs: analüütilise filosoofia seminar 20*, 259–278. Tartu: EYS Veljesto Kirjastus.
- 2010 “Kuidas saab jumal tahta, teha ja tunda?” *Akadeemia* 6, 447–472.

**Ettekanded:**

- 2013 “Läbipaistvuse läbikukkumine”, IX Eesti Filosoofia Aastakonverents, 14. detsember.  
“Second-personal attributions of mental states”, The Second-Person Perspective in Science and Humanities, Oxford, 19. juuli.  
“What to do with beliefs and desires”, Mind & Action Seminar, Radboud University Nijmegen, 15. märts.
- 2012 “The communicative role of folk psychology”, Postgrad Seminar, University of Hertfordshire, 26. november.  
“When do we need to talk about desires” [Poster], GAP.8, Konstanz, 17.–20. september.  
“First-person mindshaping” [Poster], Social Cognition, Engagement, and the Second-Person-Perspective, Cologne, 25.–27. mai.
- 2011 “The role of enactive imagination in Goldman’s simulation theory” [Poster], Understanding Other Minds – Embodied Interaction and Higher-Order Reasoning, Bochum, 20. september.  
“About the function of folk psychology”, Seventh European Congress of Analytic Philosophy, 1.–6. september.
- 2010 “Religion’s folk psychological context”, European Conference on Science and Theology, Edinburgh, 7.–11. aprill.

**Tunnustused:**

- 2013 Akadeemia Hõbeauhind artikli “Enesetõlgendus ja soovid” eest.
- 2011 Akadeemia Hõbeauhind artikli “Kuidas saab jumal tahta, teha ja tunda?” eest.

## DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS

1. **Jüri Eintalu.** The problem of induction: the presuppositions revisited. Tartu, 2001.
2. **Roomet Jakapi.** Berkeley, mysteries, and meaning: a critique of the non-cognitivist interpretation. Tartu, 2002.
3. **Endla Lõhkivi.** The sociology of scientific knowledge: a philosophical perspective. Tartu, 2002.
4. **Kadri Simm.** Benefit-sharing: an inquiry into justification. Tartu, 2005.
5. **Marek Volt.** The epistemic and logical role of definition in the evaluation of art. Tartu, 2007.
6. **Aive Pevkur.** Professional ethics: philosophy and practice. Tartu, 2011.
7. **Toomas Lott.** Plato on Belief (*doxa*) *Theaetetus* 184B–187A. Tartu, 2012, 208 p.
8. **Jaanus Sooväli.** Decision as Heresy. Tartu, 2013, 153 p.
9. **Ave Mets.** Normativity of scientific laws. Tartu, 2013, 217 p.
10. **Vivian Bohl.** How do we understand others? Beyond theories of mind-reading and interactionism. Tartu, 2014, 184 p.