

Making Sense of Response-Dependence

Eline Busck Gundersen

Submitted in application for the degree of Ph.D.

University of St. Andrews

10. November 2006

Abstract

This thesis investigates the distinction, or distinctions, between response-dependent and response-independent concepts or subject matters. I present and discuss the three most influential versions of the distinction: Crispin Wright's, Mark Johnston's, and Philip Pettit's. I argue that the versions do not compete for a single job, but that they can supplement each other, and that a system of different distinctions is more useful than a single distinction. I distinguish two main paradigms of response-dependence: response-dependence of subject matter (Johnston and Wright), and response-dependence of concepts only (Pettit). I develop Pettit's 'ethocentric' story of concept acquisition into an account of concept evolution that suggests answers to a range of hard questions about language, reality, and the relation between them. I argue that while response-dependence theses of subject matter can be motivated in very different ways, the resulting theses are less different than they might seem. I suggest that the traditional ways of distinguishing response-dependent subject matters from response-independent ones – in terms of a priori biconditionals connecting facts of the disputed class with responses in subjects in favourable conditions, and fulfilling some further conditions such as non-triviality and sometimes necessity – may not be the best approach. I also discuss two general problems for response-dependence theses: the problem of 'finkish' counterexamples, and the problem of specifying the 'favourable conditions' a priori, yet in a non-trivial way. The discussion of response-dependence is informed by a framework based on the idea that some realism disputes can be viewed as location disputes: disputes over the correct location of the disputed properties among several levels of candidate properties. The approach taken in this work is a charitable one: I try to make sense of response-dependence. The conclusion is the correspondingly optimistic one that the idea(s) of response-dependence makes sense.

Acknowledgements

I owe thanks to many people for help, inspiration, and encouragement in the process of writing this thesis. First, thanks to my excellent team of supervisors. It has been a privilege and a pleasure to work with them. Thanks to Crispin Wright for teaching me how to do philosophy, by teaching and by setting an example, and for providing more inspiration than I thought anyone could. Thanks to Katherine Hawley for essential ‘baby-sitting’ during the writing-up process, and for lots of good advice and encouragement. And thanks to Sarah Broadie for invaluable support and advice. Her support, more than anything else, helped pull me through the final stages of the process. Also, thanks to my examiners, Simon Prosser, Alex Miller and Alan Millar, for much useful feedback, and for making the viva a thoroughly pleasant experience.

Thanks to Arché, Centre for the Philosophy of Logic, Language, Mathematics and Mind, for providing an extremely stimulating academic environment. Thanks to my colleagues at the Danish research project Naturalised Mind – Cognisant Nature (Namicona) for many excellent workshops in a friendly and stimulating atmosphere. Thanks to staff and students at St. Andrews for inspiration and the rest. And thanks to Philip Pettit, Jussi Haukioja and many others for philosophical inspiration.

I am grateful to the Danish Research Education Council (Forskeruddannelsesrådet) for making this project financially possible through a full Ph.D. scholarship. Also, thanks to the School of Philosophy at St. Andrews for additional financial support. I am grateful to Department of Philosophy at Aarhus University for accommodating me during my visits to Denmark.

In a more private vein, thanks to Andy Law for – what’s the phrase? – keeping me alive and even reasonably fit, and for many valuable Ps. And for showing me that there is more to life than nerdery. Thanks to Chris Kean for hints about the importance of philosophy. Thanks to Margit for a shot of energy when it was most needed. Thanks to Sonia, Nikolaj, Katrine, Irene, Jørn, Annette, Chiara, Seba, Silvia, Dan, and all the rest of you for friendship and support through the years in St Andrews. Thanks also to my friends at home. And thanks to my family for loving and unfailing support, and for bearing with my absence and absentmindedness. Thanks to all the rest of you who have made it possible to make it this far. And finally, thanks to Lars. For everything: encouragement, curries, trips to the highlands, and for sticking with me through the long, hard years apart. This work is dedicated to Lars.

Contents

CH. 1: THE EYE OF THE BEHOLDER	1
1. Motivations for response-dependence	2
1.1 <i>Reconciling realism and subject-relatedness</i>	2
1.2 <i>A second motivation: the Wittgensteinian problem of rule-following</i>	4
2. Response-dependence and biconditionals	5
3. A short history of response-dependence	8
4. Disambiguating ‘real’ and ‘subject-related’	10
4.1 <i>Location problems</i>	15
5. Preliminaries on concepts and properties	15
5.1 <i>Referentially and criterially governed concepts</i>	16
5.2 <i>Substantial and deflationary properties</i>	17
5.3 <i>The distinction related to other distinctions</i>	19
6. Overview of the thesis	20
7. The eye of the beholder	22
CH. 2: LOCATION PROBLEMS	25
1. Location disputes	25
1.1 <i>Intuitions about colours</i>	27
1.2 <i>A diagnosis</i>	30
1.3 <i>Other examples of levels</i>	31
2. Location options	33
2.1 <i>Lower level locations (high, low, or basic physical levels)</i>	34
2.2 <i>Higher level locations (response-dependence, functionalism, phenomenal states)</i>	37
2.3 <i>How to choose a location?</i>	41
3. Alternatives to location	42
3.1 <i>Error theories: aiming too high</i>	42
3.2 <i>Primitivism: stipulating a level that has it all</i>	43
3.3 <i>Multi-level locations: functionalism</i>	44
3.4 <i>Equivocation views</i>	47
4. Summary	51
CH. 3: SOME QUESTIONS ON LEVELS	53
1. What is the relation between levels?	54
1.1 <i>The non-mereological candidate relations</i>	56
1.2 <i>Different cases, different relations</i>	61
1.3 <i>A unifying sense of levels?</i>	62
2. How to individuate levels?	64
3. Ontological differences or just modes of presentation?	66
4. Is there a basic level?	71
4.1 <i>Level fundamentalism and high-level locations</i>	74
5. The levels framework and the nature of properties	76
6. Justification and applications	79
6.1 <i>The levels framework at work for response-dependence</i>	80
7. Summary	81
8. Appendix: Theseus’s ships	83
8.1 <i>The puzzles of material constitution</i>	83
8.2 <i>Rea’s recipe for puzzles of material constitution</i>	84
8.3 <i>Level differences in puzzles of material constitution</i>	85

8.4	<i>The suggested solution in Rea's terminology</i>	88
8.5	<i>Transitivity-of-identity puzzles without level differences</i>	89
8.6	<i>Is it too easy to claim level difference?</i>	90
8.7	<i>Levels of objects or just levels of properties?</i>	91
8.8	<i>Are ships material objects?</i>	92
8.9	<i>The puzzles at work for the levels framework</i>	93
8.10	<i>Conclusion</i>	94

CH. 4: WRIGHT AND JOHNSTON ON RESPONSE-DEPENDENCE..... 95

1.	Wright: The Euthyphro Contrast	95
1.1	<i>The order of determination test</i>	99
1.2	<i>Biconditionals</i>	100
1.3	<i>The substantiality condition</i>	101
1.4	<i>Necessity and a priority</i>	102
1.5	<i>The independence condition</i>	104
1.6	<i>The extremal condition</i>	105
1.7	<i>The order of determination test</i>	106
1.8	<i>Euthyphronism about intentional states</i>	107
1.9	<i>Euthyphronism about meaning</i>	111
1.10	<i>Some objections</i>	115
2.	Johnston: Response-dispositional concepts	117
2.1	<i>Descriptive and revisionary response-dependence theses</i>	120
2.2	<i>Response-dependence and pragmatism</i>	122
2.3	<i>The Missing Explanation Argument</i>	124
3.	Similarities and differences	127

CH. 5: PETTIT'S ETHOCENTRIC STORY AND ITS POTENTIAL..... 131

1.	Pettit's ethocentric story of concept possession	131
1.1	<i>Response-dependent concepts, response-independent properties</i>	135
1.2	<i>Biconditionals and conditions</i>	137
1.3	<i>Realism and response-dependence</i>	139
1.4	<i>Some questions about Pettit's account</i>	143
2.	The ethocentric story applied to concept evolution	144
2.1	<i>Some constraints on reference</i>	146
2.2	<i>The open location question</i>	147
2.3	<i>Accessing independently existing properties by concepts made of 'for us' material</i>	150
2.4	<i>Location, intuitions, and low-level common factors</i>	152
2.5	<i>The possibility of error theories</i>	154
2.6	<i>Level-rigid and level-flexible concepts</i>	156
2.7	<i>Ethocentrism and relativist views</i>	158
3.	How are the ethocentric stories related?	162
3.1	<i>Pettit and the ethocentric story of concept evolution</i>	163
3.2	<i>Basic concepts</i>	165
4.	Summary	166

CH. 6: RESPONSE-DEPENDENCE OF SUBJECT MATTER..... 168

1.	Response-dependence theses as high level locations	169
2.	Two routes to response-dependence of subject matter	170
2.1	<i>Fitting in Johnston's contrast between descriptive and revisionary versions</i>	172
2.2	<i>Different routes; different theses?</i>	173
3.	The location dispute for dispositions	174
4.	Dispositional or subjunctive formulations?	177
5.	Substantial referents or criterially governed concepts?	179
5.1	<i>Two faces of response-dispositions</i>	180

6. Biconditionals and operationality	183
6.1 <i>Four readings of response-dependence biconditionals.....</i>	184
7. The necessity condition	185
8. The a priority condition	192
8.1 <i>A priori biconditionals and the possibility of reference failure</i>	194
8.2 <i>Flawed response-dependent concepts and a priority.....</i>	199
8.3 <i>The problem of Siberian strange-water</i>	200
8.4 <i>The parallel problem of fool's water</i>	202
8.5 <i>Strange-water-proofing strategies.....</i>	202
8.6 <i>A priori biconditionals and the a posteriori route to response-dependence.....</i>	207
8.7 <i>Summing up the discussion of the a priority condition.....</i>	208
8.8 <i>Beyond the a priority condition</i>	209
9. Location arguments as criteria of response-dependence.....	211
9.1 <i>Causal arguments</i>	211
9.2 <i>Transparency arguments</i>	213
9.3 <i>Multiple realisation arguments.....</i>	215
9.4 <i>Further problems with operationality.....</i>	215
10. Conclusion.....	218
CH. 7: FAVOURABLE CONDITIONS	222
1. Substantiality and a priority: the prima facie conflict	222
1.1 <i>Strategies for specifying C-conditions</i>	224
2. Pettit's functionalist solution	226
3. Must both requirements be met by a single specification?	229
3.1 <i>How much will be knowable a priori?.....</i>	230
4. Constancy for the wrong reasons	232
4.1 <i>Deferring to Nolan's sense of humour.....</i>	233
4.2 <i>Kappel's taste-good-drug</i>	234
4.3 <i>Wright on stable colour illusions.....</i>	235
5. Alternative job descriptions for the C-conditions	236
5.1 <i>Unruly C-conditions.....</i>	240
5.2 <i>The consequences for response-dependence accounts.....</i>	241
6. Conclusion.....	242
CH. 8: CONDITIONAL FALLACY PROBLEMS.....	245
1. Conditional fallacy problems	245
1.1 <i>Conditional fallacy problems for response-dependence theses</i>	247
1.2 <i>What is under pressure from the objections?.....</i>	250
2. Attempted solutions.....	250
2.1 <i>Johnston on response-dependence and conditional fallacy problems</i>	251
2.2 <i>Wright on response-dependence and conditional fallacy problems.....</i>	252
3. Blackburn's 'elasticity' approach.....	255
3.1 <i>Beyond the level contrast.....</i>	257
4. Relativising C-conditions?	259
4.1 <i>Some unsuccessful objections</i>	260
4.2 <i>A serious objection.....</i>	261
4.3 <i>An independence worry.....</i>	262
4.4 <i>A base-based solution to conditional fallacy problems?.....</i>	264
5. Conclusion.....	266
CONCLUSION.....	267

BIBLIOGRAPHY	273
---------------------------	------------

Ch. 1: The eye of the beholder

‘So you’re doing a Ph.D. in philosophy. What is your thesis on?’, they ask – old friends, students, distant relatives at family events. ‘Response-dependence’, I say, and watch their expression go blank. ‘Response-dependence? What on earth is response-dependence?’

This is a good question indeed. Not only because the topic is interesting, but also because there is very little agreement on the answer even among those foremost in the field. One of the aims of this work is to answer this seemingly simple question. The short answer is that response-dependence is lots of things. I shall argue that the distinction between response-dependent and –independent concepts or properties can be drawn in different ways that are of interest for a variety of purposes, and that the various versions do not compete for a single job.

I will not argue for any particular response-dependence thesis of any particular domain, but shall be concerned mainly with general issues surrounding the idea of response-dependence. Taking as starting point the three main versions of response-dependence from the literature – Philip Pettit’s, Mark Johnston’s and Crispin Wright’s – I shall discuss questions like the following: How should the distinction or distinctions be captured? What is the motivation behind them, and what is their potential with respect to which issues? What objections do they face, and how, if at all, can the objections be answered? Along the way, there will be some discussion of how to apply the distinctions to particular domains – mostly colours and mental states – but it is not within the scope of the current project to discuss such applications in a systematic way.

The discussion of response-dependence is set within a framework based on the idea of different ‘levels’ of properties in a sense to be explicated, and of certain realism disputes as disputes over the location of the disputed phenomenon among candidate properties on several different levels. I shall attempt to show how this framework provides a useful perspective on discussions about response-dependence.

In this introductory chapter, I shall start by explaining the motivations that led to the development of the response-dependence distinction(s). Then I provide an overview of some of the various versions and their similarities and differences, and a

brief history of response-dependence. Towards the end of the chapter, I shall say a bit more about the thesis, its objectives, and the approach taken to the issues under discussion. The chapter also includes a brief taxonomy of questions treated under the heading of realism, and some preliminaries on concepts and properties.

1. Motivations for response-dependence

1.1 Reconciling realism and subject-relatedness

Beauty is in the eye of the beholder, they say. What should we make of such a claim? Some properties, beauty among them, seem to be related to subjects and the way they respond to things in a way that other properties, such as shape and mass, are not. These properties are not properties of the world considered in isolation from its impact on subjects, and would presumably not figure in a completely objective and species-neutral description of the world (except perhaps in descriptions of the way certain features happen to impact beings of certain kinds). If we consider a beautiful thing in isolation from its impact on subjects, it is hard to find the beauty. We can find properties that serve as the basis of the beauty, such as shape and texture. But these properties seem the wrong kind of candidates for being the beauty itself. But if the beauty is not to be found among these properties, what is it? One fairly extreme response would be that beauty does not exist, and that all talk about beauty rests on a mistake (and similarly for whatever else would drop out of a completely objective description of the world – presumably a lot). A less extreme response would be that beauty is brought into the world by subjects and the way they experience things. A common way to cash this out is to say that beauty is a *projection* onto the objective world of the way subjects experience things. This entails a fairly strong anti-realist stance towards the subject matter in question; on this view, the disputed phenomenon has little to do with the objects to which it is ascribed, and is only visible from the perspective of certain beings. In short, it is tainted with subjectivity, and does not really count as real. But beauty seems real, despite its subject-relatedness, and despite its being different from paradigm examples of real, objective properties. Is there a way to reconcile its reality with its subject-relatedness?

Though we would like to say that beauty is a real phenomenon that subjects are in a position to appreciate, the projectivist conclusion might perhaps be acceptable in the case of beauty. But there are other domains for which this conclusion is less acceptable. Other areas, including values in general and moral values in particular, display the same combination of features: they seem to be related to subjects in much the same way as beauty is, yet we would like to think they are real. A theory that implies that they are not seems highly unsettling, and raises hard questions about how much of e.g. our moral practices could survive such a discovery. How can we reconcile realist intuitions with intuitions about a close relation to subjects?

When Mark Johnston introduced the term response-dependence in 1986¹, his project was to find an account of such domains – especially values – that could accommodate their special relation to subjects, yet vindicate a moderate realism of the properties in question. Johnston’s suggestion was that values should be understood on the model of traditional secondary qualities, and accounted for in terms of *dispositions to elicit certain responses in certain subjects in certain conditions*. His initial suggestion was that claims like

- (1) x is red iff x would look red to standard subjects in standard conditions
and
- (2) x is good iff x would seem good to standard subjects in standard conditions
can be given an interesting a priori form, as opposed to claims like
- (3) x is square iff x would seem square to standard subjects in standard conditions,
which would be true a posteriori if at all.²

¹ According to Wright (1992, p. 109, fn.16) and Johnston (1989, p. 146, fn. 8), the term was introduced by Johnston at a seminar in Princeton in 1986; it first appeared in print in Johnston (1989) (though Wright (1988) comes close).

² Wright (1992), p. 109.

In the case of moral concepts, the specification of the ‘standard’ subjects and conditions would presumably involve some idealisation; more about standard subjects and -conditions below.

Wright developed this idea into the ‘order of determination test’. The basic idea is that claims like the first two should be read right to left; x is red (,good) because it would seem red (good) to standard subjects in standard conditions. Judgements in standard conditions (partly) determine the extensions of the concepts. By contrast, claims like the third should be read left to right; square things seem square to standard subjects in standard conditions because they are square. The judgements merely reflect facts that are independently determined. This idea was developed into a much more elaborate account. We shall look at this proposal, and Johnston’s, in Ch. 4.

Accounts like Johnston’s and Wright’s allow for the subject-relatedness of the phenomenon under discussion; something falls under a response-dependent concept in virtue of the responses it would elicit from subjects. Yet the accounts also allow for the existence of facts about the matter in question, insofar as there are facts about the responses and response-dispositions in question.

Many other authors³ have explored this general paradigm for response-dependence theses: finding accounts that describe a phenomenon as somehow tied to responses in subjects, yet still real (though maybe not in all respects associated with the term, or maybe less so than response-independent phenomena). Common to the views on this paradigm is that, whether they are phrased as theses about concepts or theses about properties, they entail that not only the concepts, but also the subject matter under discussion, is dependent on, and in some sense constituted by, subjects and their responses.

1.2 A second motivation: the Wittgensteinian problem of rule-following

A second motivation for response-dependence theses is the hope that they will provide a solution to the Wittgensteinian problem of rule-following. In one version, this problem can be construed as that of explaining how a finite set of examples of applications of a rule, e.g. the rule governing the correct use of a concept, could put us on track of a rule with many more, perhaps infinitely many, possible applications.

³ Among them Wedgewood (1998) and DeClercq (2002). See also Lewis (1989), Smith (1989), Miscevic (1997), and Lopez de Sa (2003).

Given a finite set of examples, it is in principle possible to extrapolate from them in lots of different ways – most of them strange and grue-like, but nevertheless possible. But if this is so, how does a learner manage to pick out one of these as the way to go, and to pick the one that her teacher intended her to pick out?

Philip Pettit's version of response-dependence was developed as a response to this challenge. The basic idea is that the response-dispositions of subjects can pick out one way to go as the natural one, and hence play a crucial role in the way (some) concepts are acquired. The resulting account of response-dependence is very different from those versions of response-dependence that aim to capture a qualified realism that takes some form of subject-relatedness into account. Pettit's brand of response-dependence is a thesis about how certain concepts are acquired, but is neutral on the nature of their subject matter. In this, it differs strongly from accounts like Johnston's and Wright's, on which the subject matter itself is response-dependent. The two kinds of accounts address different questions, and have different implications regarding many questions about realism and related issues.⁴

2. Response-dependence and biconditionals

As the name suggests, it is agreed on all hands that response-dependent concepts somehow involve certain responses in certain subjects and conditions. Also, advocates of different versions of the idea tend to agree on some basic conditions that a concept must meet in order to count as response-dependent. More precisely, almost everyone⁵ agrees that response-dependent concepts underwrite a priori true biconditionals of roughly the form

- (4) x is F (i.e. something, x , falls under the concept 'F' or has the property F) \leftrightarrow x would elicit response R from subjects S in circumstances C,

⁴ However, Pettit's version of response-dependence implies that one particular subject matter – rule-following – is subject-related yet real, and for this domain, his view and its motivations resembles those of other authors on response-dependence. Crispin Wright develops a similar view in his (1989); more on this in Ch. 4. So the second motivation for response-dependence claims that I have distinguished can be seen as a specific instance of the first.

⁵ Though see Miscevic (1998) for an argument against the apriority of the equations.

where R, S and C can be specified without trivialising ‘whatever it takes’-clauses. Assuming for a moment that colour concepts are response-dependent, an example could be this:

- (5) X is red iff x would elicit the judgement that x is red (or the belief that x is red, or the relevant colour experience) in actually statistically typical humans who are competent with colour concepts, in conditions of normal daylight, proper distance to the object etc.

(For now, let’s bracket any worries about this particular specification, such as what ‘proper distance’ means and whether it is a whatever-it-takes-clause.)

This general recipe provides wide scope for variation. First, the responses, subjects and conditions can be specified in very different ways. The responses could be cognitive responses like judgements (these could be either judgements that x is F or other judgements), or phenomenal responses such as colour experiences or pains, or, on some accounts, a combination of different responses. Purely bodily responses, such as nausea, sweating, digesting etc. are normally excluded.

The subjects and conditions could be normal or idealised, statistically typical or actually statistically typical, etc. Or they could be specified fully in physical and physiological terms. Depending on these specifications, especially on whether the specifications are flexible or rigidified (e.g. ‘statistically typical conditions’ versus ‘*actually* statistically typical conditions’), the resulting theses will be very different.⁶ One choice can make an equation false while another will make it true, so objections against one choice of equation for a particular domain is not an objection to response-dependence theses for that domain in general.

Throughout this work, I shall use the term ‘C-conditions’ as a shorthand for conditions that play this role. It is useful to have such a general term because it is often better left open whether conditions needed are normal or ideal ones (e.g. for generalisations across different domains with different sorts of C-conditions), and because it makes clear that it’s this technical use I am after. Pettit uses the term

⁶ According to Mark Lebar (2005), it is agreed on all hands that the specifications must be rigidified. But many authors, including Wright, Johnston and Pettit, allow for flexible components in the specifications of R, S and C.

favourable conditions; other common terms are standard, normal, and ideal conditions, but these are too specific to be applicable to all cases. Often, the conditions on the subjects are included in the specification of the C-conditions; where nothing hinges on separating them, I shall follow this convention.

The way the claim ‘x is F’ is to be understood also provides scope for variation. Some authors (including Wright and Pettit) present response-dependence as a property of *concepts*, and the biconditionals say what it takes for instances to fall within the extension of the concepts. Other authors (e.g. Wedgwood and Lopez de Sa) take response-dependence to be a property of *properties, subject matters or states of affairs*, and so take the biconditional to be saying something about these rather than about concepts. (As we shall see, Wright’s version is naturally interpreted as response-dependence of subject matter, despite being presented as a thesis about concepts.)

Many authors impose further conditions on the equations that must be fulfilled for a concept to qualify as response-dependent (apart from the conditions of a priority and substantial specifications of R, S and C mentioned above). For example that the biconditional must be necessary (Johnston), or that the specification of R, S and C must presuppose nothing about the extension of the disputed concepts (Wright), or that there must be no better explanation than response-dependence of the fact that the biconditional holds and the other conditions are fulfilled (Wright). We shall examine the motivation for these conditions in Ch. 4.

Another point of variation is that some authors prefer overtly dispositional formulations, replacing ‘would’ with ‘is disposed to’. For reasons to do with conditional fallacy problems (see Ch. 8), many, including Johnston and Wright, take the two to be interestingly different.

Finally, as we have already seen, there are surprisingly large differences in the philosophical content the distinction between response-dependent and –independent concepts is meant to capture, in the problems it is meant to solve, and also in the implications of response-dependence theses for the level of realism in the disputed domain. All this will be described in more detail in Chs. 4-6. This brief introduction to response-dependence should suffice merely to give an idea of what we are talking about.

3. A short history of response-dependence

The idea that some concepts and properties are in some way dependent on responses in subjects is not new. In fact, it can be traced back to antiquity; Plato's dialogue *Euthyphro* contains an example of a response-dependence view, or something very close (see Ch. 4). Other classical discussions on the subject are found in Galileo and Hume, who both take a projectivist view on traditional secondary qualities like colours, smells, tastes etc.⁷ Most importantly, perhaps, it is found in Locke's distinction between primary and secondary qualities, secondary qualities being qualities that 'in truth are nothing in the objects themselves but powers to produce various sensations in us'.⁸ This distinction inspired the response-dependence distinction to such an extent that many philosophers use the expressions 'response-dependence views' and 'secondary quality views' interchangeably.

The primary/ secondary quality distinction was reintroduced into contemporary discussions on realism and its opposites by John McDowell and David Wiggins.⁹ They explored the connections between values and responses in subjects in the hope of vindicating a moderate realism about values which was compatible with their subject-relatedness. Johnston coined the term 'response-dependence' while continuing this project.

The first wave of writings on response-dependence in the late 1980's had as one of its main objectives the search for an account of values that would vindicate a moderate moral realism while acknowledging the dependence of values on human response-dispositions.¹⁰ The optimism about this project was halted somewhat by Simon Blackburn's paper 'Circles, Finks, Smells and Biconditionals', which argues that such attempts face insurmountable difficulties in specifying the relevant responses, subjects and conditions in a satisfactory way. If concepts of the disputed kind are employed in the specifications, it is questionable whether the biconditionals can do the work they were intended to do; for example, they will not be helpful as explanations of the concepts to someone who doesn't know them. If the

⁷ Galileo (1957), Hume (1975) and (1978).

⁸ Locke (1951), p. 71 (II, viii, 10).

⁹ McDowell (1981) and (1985), Wiggins (1987).

¹⁰ See Wright (1988), Johnston (1989), Smith (1989), and Lewis (1989).

specifications are made in purely naturalistic terms, it can be argued that the equations will be a posteriori true at best, and often false.¹¹

Another recurrent theme in the early period was the project of using response-dependence to make progress with the Wittgensteinian problem of rule-following. Wright and Pettit both developed proposals of this kind.¹²

The main versions of the response-dependence distinction – Johnston’s, Wright’s, and Pettit’s, were in play within the first five years of the introduction of the term ‘response-dependence’, though some of the work took some years to appear in print.¹³

Most of the main objections to the idea of response-dependence also came up in the very beginning. One is Blackburn’s dilemma which was mentioned above. Another is Mark Johnston’s ‘Missing Explanation Argument’.¹⁴ Johnston argues that response-dependence theses are in tension with causal explanations of responses in terms of facts of the disputed class. Put very crudely, if something is red because it looks red, it can’t also be the case that it looks red because it is red. However, in many domains where a response-dependence view is an option, explanations of the latter kind make sense. The inability to accommodate such explanations thus counts against response-dependence views of those domains.

The next wave of work displayed a lot of variety, both in the attempts to cash out the distinction(s) between response-dependent and –independent concepts or properties (see e.g. Wedgwood 1998 and DeClercq 2002; Pettit’s line was continued and developed by Haukioja 2000), and in the wide range of application areas where the potential of response-dependence theses were explored. Examples include colours (e.g. Johnston 1992, Wright 1989, 1998 etc., and dispositional theories of colours in general), moral values (Lewis 1989, Smith 1989, 1994, 1998, Johnston 1989, Wright 1988, Lopez De Sa 2003), aesthetic values (Zangwill 2001), modality

¹¹ Blackburn (1993), p. 270-74.

¹² Wright (1989), Pettit (1990).

¹³ Core sources include Pettit (1990) and (1991), Johnston (1992) and (1993; publication somewhat delayed), and Wright (1989) and the appendix on response-dependence in (1992), which is based on the widely circulated manuscript ‘Notes on Basic Equations’.

¹⁴ Johnston (1991) and (1998). The argument is presented in a bit more detail in Ch. 4.

(Menziez 1998), abstract objects/mathematics (Divers & Miller 1999), rule-following/ meaning (Pettit 1990, Wright 1989), and social institutions (Hindriks 2005).

The present situation is that there is no consensus about what exactly it is for a concept or property to be response-dependent, and very little consensus about what concepts or properties are likely to be response-dependent (even for reasonably well defined candidate versions of the response-dependence distinction). But at least there is a growing awareness that different authors mean different things by response-dependence, and that this had better be taken into account when assessing or applying their ideas.¹⁵ One of the aims of this work is to contribute to the awareness of these differences, and to warn against the misunderstandings likely to result from the ‘one name, many ideas’ situation.¹⁶

4. Disambiguating ‘real’ and ‘subject-related’

Before we proceed to the discussion of response-dependence, there are a few preliminaries we need to get in place. The first concerns the notions of ‘real’ and ‘subject-related’ which is crucial to the motivation of response-dependence in most of its forms. I have been using the terms ‘subject-related’ and ‘real’ as if everyone knows what they mean. But both are notoriously slippery, and can be understood in many different ways. Distinguishing these will be useful in the following.

‘*Subject-related*’ can mean at least the following things:

- 1) that the properties in question are *relational rather than monadic* – that they are relations involving subjects, rather than properties of the entities or whatever to which they are normally ascribed.
- 2) that the properties (or states of affairs) in question are properties (/states of affairs) *of subjects or of systems that include subjects as parts*.
- 3) that the properties are *dispositions to elicit certain responses in subjects*.

¹⁵ See e.g. Pettit (2006).

¹⁶ For what appears to be an example of such misunderstandings, see Devitt (2006).

- 4) that the properties *exist in virtue of our classifications*, and are not properties in the substantial sense to be defined in the next section. (Or, for states of affairs, that these fall under a certain concept in virtue of our classifying them as such).
- 5) that the properties or states of affairs *are in some (other) sense brought about by subjects* (causally or similar) – that the states of affairs come to be because of certain features or actions of or facts about subjects – e.g. that they are created by projection of the experiences of individuals, or by consensus. (This option is different from 4; for example, an object in a subject-free environment or world could be subject-related in the sense of 4, but not 5. This option is also fairly uninteresting in the context of response-dependence).

Which of these senses is relevant in the context of response-dependence? This is one of the issues that we shall investigate. There may be many different answers, applicable to different kinds of concepts and corresponding to different ways of drawing the distinction or distinctions between response-dependent and independent concepts and subject matters.

‘Real’ is an even more equivocal term. Positions labelled ‘realism’ get the label in virtue of giving affirmative answers to some or all of a range of questions to be distinguished below. Negative answers to the various questions yield very different positions. ‘Anti-realism’ as a label for all (or even some loosely defined group) of them is thus not a very helpful label.

Realism is not a question of all or nothing. First, you can combine positive answers to some of the questions with negative answers to others. Secondly, you can hold different combinations for different domains; you can be an anti-realist (in a given sense) about unicorns without being one about electrons, or vice versa. The main questions discussed under the heading of realism are these:

- 1) a semantic realism question: Are sentences of the disputed kind apt for truth or falsity?

Views that return a negative answer to this question include emotivism and prescriptivism in ethics¹⁷ and various instrumentalist positions in the philosophy of

¹⁷ Ayer (1936), Hare (1952) and (1991).

science. Common to them is that they ascribe the sentences in question some other role than stating truths (voicing emotions, prescribing attitudes or actions, etc.). Response-dependence views uniformly return an affirmative answer to this question. Secondly, there is

- 2) an epistemic realism question: in most or core cases, do we get things right regarding the properties, entities, or states of affairs in question?

A negative answer amounts to scepticism about the domain in question. The answer given by response-dependence theorists will be that in C-conditions, we do get response-dependent matters right. (However, this does not amount to a counter-intuitive infallibility in a given case, as we never get a guarantee that we are in fact in C-conditions.) A third realism question is

- 3) an ontological realism question: do the entities/ properties/ states of affairs in question exist independently of us?

This question is a combination of several questions, among them those that are most relevant in the context of response-dependence. First, we should distinguish these two questions:

- 3a) Do the entities/ properties/ states of affairs exist?

A negative answer is given by error theorists of the domain in question, e.g. Mackie on ethics, Field on mathematics, Boghossian and Velleman on colours, and most of us on phlogiston.¹⁸ Response-dependence theorists about a domain will give an affirmative answer (provided there are things that would elicit the relevant responses in C-conditions). The second and crucial question is this:

- 3b) Do they exist independently of us?

This question calls for further disambiguation. In effect, it is the question about subject-relatedness, for which we have already distinguished a number of meanings. Here, we shall distinguish a few more, based on different disambiguations of the formulation in b).

¹⁸ Mackie (1977), Field (1989), Boghossian & Velleman (1989).

First, there is scope for variation with respect to the worlds or situations in which subjects might be required for the phenomenon to exist. A fairly lean sort of realism in this respect would be entailed by the claim that

- Entities/ properties/ states of affairs of the disputed kind (let's call them Fs) could exist or be instantiated *even if there were no subjects present to appreciate (perceive etc.) them.*

A denial of this claim amounts to strong, Berkeley-style idealism, and to a variety of subject-relatedness of kind 5) as discussed above. More substantial realist claims would be

- that Fs could exist even *in a world where no-one would ever be around to appreciate it* (denial: response-dependence views that take responses of the subjects in the world in question as the relevant ones), or
- that Fs could exist *even if there were no actual subjects to set the standard according to which the world under consideration can be evaluated*, i.e. that it exists independently of our classifications (denial: other response-dependence views; views that the property exists only in virtue of classifications, i.e. that it is subject-related in the sense of 4 above)

Secondly, there is scope for variation with respect to *the features or responses of subjects that Fs are relative to*. Here are some options; there are many more:

- Independence of individual judgements in particular cases?
- Independence of judgements of entire communities?
- Independence of best judgements?
- Independence of best judgements of entire communities?

The choices made regarding these questions (and how 'best judgements' are specified) will also have implications for the level of realism regarding Fs.

Thirdly, claims about existence independently of subjects could also be taken to concern the sorts of things that the property is a property of (subject-relatedness in sense 2) above):

- Are the Fs properties (/states of affairs) of entities or systems that do not involve subjects?

This could be denied by claiming that they were *properties of subjects*. This position would be taken by projectivists and, uninterestingly, by common-sense views of properties that are standardly ascribed to subjects, e.g. being a child, a truck driver, in pain, or 1.75 m. tall. It would also be denied by people who hold that the properties or states of affairs concern *a system that involves subjects as a part*, e.g. views that identify the properties or states of affairs with response-dispositions of subjects.

Besides these options, there are the other versions distinguished under subject-dependence above. We might distinguish more ways to deny 3b (that Fs exist independently of subjects) – at least as many as there are versions of response-dependence. The project of capturing the response-dependence distinction could be described as that of finding the right way to deny 3b) while giving affirmative answers to questions 1 (truth-aptness), 2 (epistemic access) and 3a (existence of Fs). As already suggested, there may be several interesting ways to do this, and corresponding different versions of response-dependence.

This overview is not exhaustive, and is only meant to communicate a sense of the complexity of the issues treated under the heading of realism. We shall devote more attention to these issues, in particular the different ways of denying 3b), in Chs. 4-6.

The different realism questions are connected in various ways. An important connection is described by Wright:¹⁹ A modest ontology (i.e. a realist answer to (some relevant interpretation of) 3b), the question about subject-relatedness) goes with a presumptuous epistemology; if Fs are independent of us and our capacity to recognise them, there is a substantial epistemic achievement involved in detecting them, but this opens the possibility that we might always get it wrong. The view thus invites scepticism. One way to respond to such worries is to move the target nearer, and claim that our capacities are somehow involved in constituting the Fs in question, in which case we get a modest epistemology – it takes less of an achievement to get things right regarding the Fs – but at the price of a contentious ontology (the F's don't exist independently of us in the way we might have expected).

¹⁹ Wright 1992, p. 1-7.

Response-dependence theses of different kinds might be seen as attempts to find the best compromise between these two types of considerations in a given domain. For each domain, we should seek the version that strikes the best balance between realist intuitions regarding ontology and realist intuitions regarding our epistemic access to the phenomena in question. For many domains, this balance might be best achieved by giving a negative answer to some version of question 3b) about subject-relatedness. This is what makes the idea of response-dependence attractive. For different domains, it may be different interpretations (/different ways of denying 3b) that offer the compromise that best preserves intuitions about the subject matter in question. This is what makes it attractive to operate with a system of different distinctions rather than choosing one of them as the ‘right’ way to draw the response-dependence distinction.

4.1 Location problems

Some of the disputes nominally about realism or lack of same with respect to a domain of Fs can fruitfully be viewed as location disputes – disputes about which of a range of candidate properties is the referent of ‘F’, or on what level (in a sense to be explicated) the Fs should be located. (In the case of colours, for example, the candidates are surface structures, reflectance spectra, colour experiences, and dispositions to elicit them.) The location issue has implications for the realism questions regarding Fs because different locations will yield different answers to some realism questions. In Chs. 2 and 3, I shall develop a framework based on the idea of levels and of realism disputes as location disputes. This framework will serve as a backdrop for the discussion of response-dependence, and helps to bring a range of important issues into focus. It also allows us to distinguish clearly the various versions of response-dependence, and to bring out their potential (especially that of Pettit’s suggestion, which is the subject of Ch. 5) and implications.

5. Preliminaries on concepts and properties

Before we proceed to investigate this framework, some further preliminaries are needed. They concern two ways a concept can function and two ways to think about properties.

5.1 Referentially and criterially governed concepts

Following Wright (2002), we can distinguish two ways a concept might function. A concept might be intended to refer to a substantial property (a property that has an extension independently of the way any concept for it works), and to have its extension determined by the extension of that property. Alternatively, the concept might be criterially governed, i.e. apply to anything that satisfies its application conditions, whatever its nature might be.

Let us look at the two options in a bit more detail. First, the concept might be intended to latch on to a substantial property – paradigmatically, a natural kind. If the concept succeeds in picking out such a property, the extension of that property will determine which things fall within the extension of the concept. The concept of water presumably works in this way: the intuitive content associated with the concept – the ‘water symptoms’ (clear, tasteless, odourless liquid found in oceans and rivers which quenches thirst etc.) – point us to a natural kind, H_2O , which is common to most of the samples that display the water symptoms, and which is responsible for these symptoms. Once this property is recognised as the referent of ‘water’, the water symptoms have finished their job, except as a rough and ready guide to the extension. The detail of the extension of ‘water’ is determined by the property of being H_2O , not by the water symptoms; something that displays the symptoms but is not H_2O is not water, while H_2O that for some reason fails to display the symptoms (e.g. by being polluted by other substances) is water. We can call such concepts referentially governed concepts, as the important thing about them is reference to a (independently given) property.

Criterially governed concepts, by contrast, are concepts for which a set of criteria do the extension-determining. In order for something to fall under a criterially governed concept, all that is required is that it satisfies the criteria. Even if ‘water’ fits the paradigm of referential concepts, there could have been a concept, ‘swater’, that worked in this way. For something to be swater, all that would be needed is that it displays the water symptoms; every sample of clear, tasteless, odourless liquid etc. would be swater, even if they were not H_2O , whereas instances of H_2O that did not meet these criteria (e.g. due to pollution) would not be swater. Likely examples of criterially governed concepts would be ‘constitutional’ – to be

such that the relevant body (highest court or whatever) does not consider it unconstitutional – and ‘pain’; intuitively, if something hurts, it’s a pain, regardless of its underlying nature.

As with most of the distinctions drawn in this work, this distinction is clear enough in principle, but less clear-cut in its applications to concepts of natural language. Thus, real-life concepts may equivocate between a referentially and a criterially governed version, containing material or having uses that suggests both. Or intuitions connected with the concept may simply be silent on whether or not it is intended to refer to a substantial property, leaving the concept open to precisification both ways. Also, intuitions may be adjusted, and concepts might change. It might happen that we’ve been expecting a concept to refer to a low level property and function as a referentially governed concept, and that it turns out that no suitable referent is to be found, but that the concept has enough uses that survive this discovery to warrant the conclusion that it is really a criterially governed concept. Conversely, perhaps a seemingly criterially governed concept may turn out to be correlated with a property that can fruitfully be viewed as its referent, e.g. if this makes for a more consistent use of the concept than going by the criteria (perhaps because the criteria are flawed in some way, by inconsistencies or by failure to settle the use of the concept in many and important cases). But despite these complications that arise for particular applications, in principle the distinction is clear enough. Or at least it is clear to the same extent as the notion of *property* used in the explication. Let us try to explicate that notion.

5.2 Substantial and deflationary properties

For the purposes of our discussion, we can remain neutral on many of the hard issues surrounding the metaphysics of properties. However, we need to distinguish between two conceptions of properties that we shall label deflationary and substantial (also following Wright). On the deflationary conception of properties, instances need have nothing in common apart from their being classified together under a concept. In the words of an adherent of this view of properties, Simon Blackburn, properties are

nothing more than ‘the semantic shadows of predicates’.²⁰ On the substantial conception of properties, instances of a property must have more in common than this: they must have something in common in virtue of which they could determine (the detail of) the extension of a referentially governed concept – an extension that would be independent of our classifications, which, if we are lucky, may succeed in tracking such properties and ‘carving reality at its joints’. (Of the main players in the field of response-dependence, Pettit seems to endorse such a view. So, presumably, does Johnston.)

One may choose one or other of these views across the board. One might think that properties must be properties in the substantial sense in order to be properties at all, and that ascriptions of merely deflationary properties is just a roundabout way of saying that there is no property there. Alternatively, one might think that there are properties only in the deflationary sense, and that looking for classifications made right by things in themselves is a misunderstanding. Blackburn (1993) endorses such a view, and draws just the distinction we are after:

I myself see properties as the semantic shadows of predicates, not as self-standing objects of investigation. But others see matters differently: they think of there being a substantive science of identifying the property ϕ which is not the same thing as identifying the concept ϕ , or the role ϕ judgements play in our thought. [...] Property identity becomes a topic of its own. I shall call this the substantive way of thinking of properties. (Blackburn 1993, p. 262)

Apart from these two ‘pure’ views on the nature of properties, there is also a third option: a mixed view that allows for properties of both kinds. On such a view, some properties may be substantial, natural properties, while others may be properties merely in the deflationary sense. (Such a view would be a natural choice for Wright, though he might also stick with a pure deflationist view.) A variation of the view might even allow that substantiality is a matter of degree – that there are more or less substantial properties as well as substantial and non-substantial ones. (In Ch. 3, I shall argue that such a view is natural given the framework I develop in Chs. 2 and 3.) Such differences in degree of substantiality may be explained in several ways.

²⁰ Blackburn 1993, p. 262; quoted below.

They may arise because the common factors between instances of the properties are of different degrees of importance, or of different extent, or because the common factor is recognisable from more or less species-specific standpoints. A lot of refining could be done here, but we shall stick with a relatively simple picture for now.

5.3 The distinction related to other distinctions

How does the distinction between substantial and deflationary properties relate to 1) the question of realism versus nominalism about universals, and 2) the distinction between sparse and abundant properties?

With respect to the first question, it is natural to combine belief in deflationary properties with nominalism about universals.²¹ Conversely, it is natural to combine belief in universals with a view of the corresponding properties as substantial. But neither combination is forced. Though the idea of deflationary properties is naturally paired with nominalism, one might take the view that for each concept (actual and perhaps even possible), there is a universal uniting the things that fall within that concept's extension (this view is weird, but the issue here is its possibility, not its plausibility).²² Conversely, you might believe that things that fall within the extension of some concepts are united by similarities of nature without believing that what unites them is a universal.

With respect to the second question, the belief in deflationary properties naturally goes with a view that properties are abundant, and that any possible classification, natural or otherwise, corresponds to a property. (However, in principle, it would be possible to stop short of the belief that any set of individuals corresponds to a (deflationary, abundant) property; one might hold that it takes a concept – actual or maybe just possible – to bring a property into existence, though this would be a strange and strongly anti-realist view.) A substantial view of

²¹ 'Nominalism', like 'realism', is a name used for many different positions; it can also be understood in a sense akin to the deflationary view of properties.

²² This would require that this universal does not count as a 'further similarity in virtue of their nature' uniting the instances of the property. This requirement can easily be met; if it doesn't seem natural (as it does to me), we can adjust the definition of substantiality to accommodate it.

properties naturally combines with the view that properties are sparse. Belief in substantial properties can be combined with belief in abundant ones as well, however, provided one takes a mixed view of properties and allows for both substantial and deflationary, and sparse and abundant, properties. Also, you could presumably believe in sparse properties without believing that what unites them is substantiality in the sense described; you might require, in addition, that they must be ‘fundamental building blocks of nature’ or similar, whereas less might be required for substantiality. These matters deserve further attention, but these few remarks will suffice for our purposes; where related issues come up in the following, they will be discussed as needed.

6. Overview of the thesis

I shall conclude this introductory chapter with an overview of the thesis, and some words about the approach taken in this work. The thesis is in three parts. The first elaborates the framework just mentioned (Chs. 2-3). In the second, I present, distinguish, and discuss various versions of response-dependence (Chs. 4-6). The third (Chs. 7-8) treats two particular problems that arise for response-dependence theses. Here are the contents in a little more detail:

Ch. 2 introduces the idea that some realism disputes can be viewed as location disputes: disputes over the correct location of the disputed properties (or objects, or states of affairs) among several levels of candidate properties. The chapter provides an overview of standard location options – physicalism, response-dependence theses, functionalism, subjectivism, etc. – and standard arguments in location disputes. I also outline some alternatives to choosing a location, including eliminativism, primitivism, and multi-level location.

This framework raises a lot of questions: What is the relation between levels and the properties that occupy them? How to individuate levels? Are the levels ontologically different, or do they just represent different modes of presentation? Is there a basic level? Does the framework presuppose a particular metaphysics of properties? These questions, and others like them, are the subject of chapter 3.

In chapters 4 to 6, I present and discuss the main versions of response-dependence. The overall divide is between Pettit’s version, which is neutral on the

location issue and the nature of the disputed properties, and concerns only the way concepts are acquired, and versions like Wright's and Johnston's that make response-dependence a property not only of concepts, but also of their subject matters. Wright's and Johnston's versions of response-dependence are presented in Ch. 4. For each version, I present the intuitive content, formal apparatus, and intended applications, and discuss some of the problems the accounts might raise. I summarise the most important similarities and differences between the accounts, drawing in other authors where appropriate.

Chapter 5 concerns Pettit's very different response-dependence proposal. The first half of the chapter presents Pettit's views, most notably his 'ethocentric' account of concept acquisition. In the second part of the chapter, I apply this story about the concept acquisition of individuals to the concept acquisition of entire communities. This application promises answers to some hard questions about concepts, properties, linguistic practices, and the relation between them.

Chapter 6, like Ch. 4, concerns response-dependence of subject matter. I distinguish two ways in which a conclusion of strong response-dependence may be motivated, and relate them to the distinctions from the literature. I also discuss two points of disagreement between Johnston and Wright: the question whether response-dependence theses should be formulated in dispositional terms or in terms of subjunctive conditionals, and the question whether response-dependent concepts are criterially or referentially governed. I argue that both disagreements are less substantial than they seem. Finally, I discuss some of the conditions – a priority and necessity – that are often used in making response-dependence distinctions operational (i.e. in making them helpful in determining the status of a domain we are not sure about), and suggest that there are better ways to make the distinction operational than via these conditions.

Chapters 7-8 concern two challenges to the idea of response-dependence. Chapter 7 addresses a worry about the C-conditions. Advocates of response-dependence uniformly require that the C-conditions must be specifiable in a substantial, non-trivial way, yet a priori. But these requirements seem *prima facie* conflicting. I consider a functionalist solution proposed by Pettit, apply it to

response-dependence of subject matter, and discuss the problems and prospects for different ways of filling it out.

The topic of chapter 8 is ‘conditional fallacy’ problems, familiar from the conditional analysis of dispositions. I discuss what such examples show and how they may be resisted. I examine a range of defence strategies, and conclude that the counterexamples pose no serious threat to response-dependence theories.

Finally, I sum up the results and outline some perspectives and topics for further investigation. I conclude that response-dependence theories in different versions have lots of important work to do.

I leave out many relevant issues. First and foremost, I do not attempt to decide which domains are response-dependent in what senses. Also, though I mention them in passing, some of the objections to response-dependence views are not discussed in nearly the detail they deserve (Johnston’s missing explanation argument, circularity worries, worries about globalisation of response-dependence theses for meaning, etc.). I also say relatively little about the relationship between response-dependent properties and their physical correlates. Finally, there are a lot of issues that are highly relevant, but must be left out simply because they are too big, including questions about the metaphysics of properties, essences, a priority, necessity, reference, etc.

7. The eye of the beholder

My approach in this work is a charitable one: I try to make sense of – and give sense to – the idea, or ideas, of response-dependence. In this, I differ from many; apparently, response-dependence is a subject that invites people to form a very definite view on little acquaintance. More often than not, that view is that a knock-down objection makes the idea unworthy of further investigation. However, there is remarkably little consensus on what the knock-down objection is. There are a number of objections on the table, among them Johnston’s missing explanation argument, Blackburn’s dilemma concerning the specifications of the C-conditions etc., (other) worries about circularity, and the problems treated in chapters 7-8. Each is seen as THE knockdown objection by some, and as easily answerable by others (many of whom locate the knock-down objection elsewhere, where those who take

the first objection very seriously see only a small problem which can be easily addressed).²³ Philosophers being what they are, this pattern is not uncommon. However, it suggests caution about dismissing the idea on the basis of any apparent knock-down objection.

In this work, I will take the idea of response-dependence as seriously as possible and investigate its potential mileage. This approach is, of course, motivated by a hunch that there is something important in the idea, even if it turns out that we have no completely satisfactory formulation of it at present. (This goes for both overall versions – response-dependence of subject matter, aimed at reconciling realist and subjectivist intuitions about the domain in question, and response-dependence of concepts, aimed at solving the Wittgensteinian rule-following problem.)

While I hope to bring out some of the potential of response-dependence views, the investigation is unavoidably inconclusive. Some hard issues are left in an inconclusive state; while some sort of answers to the objections are always forthcoming, they are often less than completely satisfactory. Consequently, the proper conclusion to draw from the discussion will often depend on preferences already in place. If you are willing to give response-dependence a chance, I can show you some of the rewards in view if we can make the idea work, and offer some suggestions on how to do so. If you are hostile to the idea of response-dependence, I can show that there are answers to give to your objections, but it will be hard to make the case convincing enough for people who have strong preferences against it. I presuppose a lot, including the soundness of the level-based framework to be presented in Ch. 2 and 3, and that things can be real even if they don't resemble paradigm examples of independence of a subjective standpoint.

In short, I invite the reader to give the idea of response-dependence a chance, and see what results. In this respect, philosophical views resemble children, philosophy students, and people in general: Give them a chance, and they may give much in return; meet them with suspicion, and they will live up to your worst fears.

²³ An example of such a disagreement is found in Johnston (1993) and Blackburn (1994). Johnston takes the missing explanation worry very seriously, but thinks that circularity worries can be easily addressed. Blackburn, on the other hand, thinks the missing explanation argument can be deflected, but takes circularity considerations to be fatal to the proposal.

That's a different sense of 'response-dependence', which I won't say much about it in this work, but which is very familiar from everyday life; a lot depends on the eye of the beholder.

Ch. 2: Location problems

In this chapter, I shall introduce the idea of location disputes. The core idea is that some realism disputes can be viewed as disputes over the ‘location’ of the disputed phenomenon at one of several different levels of properties. Colours will be my main example of such disputes; other domains where the approach might be useful include heat and other traditional secondary qualities, mental states, and dispositions. I shall outline some general location options, and some common argumentative strategies pointing to different locations. I shall also discuss some alternatives to choosing one of the candidate properties as the correct location.

The idea of different levels of properties raises a lot of hard questions: How are levels in the suggested sense to be individuated? What are the relationships between neighbouring levels? Do the levels represent ontological differences, or just different modes of presentation of the same phenomenon? Is there a basic level? What notion of properties do we need for the levels framework to make sense? What is this framework good for, and how can it be justified? These questions will be the subject of chapter 3. For the purposes of this chapter, we shall rely on an intuitive understanding of the idea of levels (which tends to make sense to most people given only a rough sketch and some examples).

1. Location disputes

In ‘From Metaphysics to Ethics’, Frank Jackson introduces the realism dispute about colours as follows:

There is an important sense in which we know the live possibilities as far as colour is concerned. We know that objects have dispositions to look one or another colour, that they have dispositions to modify incident and transmitted light in ways that underlie their dispositions to look one or another colour, that they have physical properties that are responsible for both these dispositions, and that subjects have experiences as of things looking one or another colour. We also know that this list includes all the possibly relevant properties. [...] Colour thus presents a classic example of the location problem. The colours must, if they are instantiated anywhere, be findable somehow, somewhere in accounts that mention dispositions to look coloured and affect light, the

physical bases of these dispositions, and colour experiences; it must be the case that some of these properties have colour names as well as names from our list. Our question is, which ones? (Jackson 1998, p. 87)

Jackson's suggestion is that the dispute about the nature of colours should be viewed as a dispute about which of a range of relevant properties is the correct location of the colours. Such a dispute might seem a quibble over a word rather than over issues that would increase our knowledge of the world. But viewed from another angle, it is a substantial dispute about what kind of properties colours are, how objective they are, whether they are physical properties or something else, whether and how they are related to subjects, what systems they are properties of, etc. Different locations will have very different implications with respect to these and other questions.

As Jackson points out, we can distinguish at least four different levels of features that are all *prima facie* relevant to colours: the surface structures of objects, dispositions to reflect light with certain distributions of wavelengths, dispositions to elicit colour experiences in suitable observers, and colour experiences themselves. Participants in the location dispute about colours generally agree that these factors are all relevant to colours. The disagreement concerns the question which one of them (if any) is most essential to colours, or on which level the reference of colour concepts should be located.¹

In rough outline, the positions corresponding to the various location options can be summarised as follows. Physicalists about colour locate the reference with physical properties. Some physicalists, including Jackson,² locate them with surface structures of objects (categorical physical properties). Other physicalists, e.g. David Lewis,³ hold that colours are dispositions to reflect light with certain distributions of wavelengths. Subjectivists about colour locate the colours with properties of subjects which are projected onto the environment – colour experiences, properties of our visual field, or similar. This line of thought is usually combined with an element of

¹ These are different questions. They will normally, but not always, be answered together. (See sect. 2.2 for an exception.)

² Jackson (1998).

³ Lewis (1997b).

error theory; see e.g. Boghossian & Velleman.⁴ Dispositionalists, e.g. Johnston,⁵ take an intermediate path and locate the colours with dispositions to elicit colour experiences in appropriate subjects. A different but related option is a Wright-style response-dependence view.⁶ Dispositionalist and subjectivist positions both fall under the heading secondary quality views (though some reserve it for one or other of these positions).

Other positions do not consist in a straightforward choice of location. One is primitivism, the view that colours are simple *sui generis* qualities that cannot be reduced to any of the associated properties, but display most or all of the features we associate with colours. Another option is eliminativism, the view that the whole practice of colour ascriptions is based on a mistake, and should either be abandoned altogether or at best accepted as a sort of figurative speech projecting subjective experiences onto the world. A third option is to recognise that the concepts equivocate between several levels, and leave the matter there, refusing to single out one level as the all-important one. A fourth option is to argue for a multi-level location, e.g. by telling a functionalist story about roles and realiser properties, where both are relevant to the phenomenon in question. We shall look at all these options in the following sections. First, we shall examine the fairly straightforward location options, and later the alternatives to straightforward single-level location.

1.1 Intuitions about colours

How do we choose between the location options in a location dispute like the one about colours? There are two groups of 'data' to measure the locations against: intuitions about colours and scientific facts about colours. I won't say much about the latter, but merely note that it seems an indisputably bad idea to choose a theory of colours that flies in the face of well-established scientific data about them. The status of common-sense intuitions is perhaps more controversial. While some philosophers hold that intuitions are what give colour concepts their content, others might argue that they are irrelevant. Externalists about meaning might hold that the question what

⁴ Boghossian & Velleman (1989).

⁵ Johnston (1992).

⁶ E.g. Wright 1989; much more on this and its relation to Johnston's views in Chs. 4 and 6.

properties the colour terms refer to depends only on certain causal connections (or whatever properties your favourite externalist account centres on), and that what competent users of these terms happen to think about colours is a completely independent, and hence irrelevant, matter. Even if the latter is the case, though, it would seem that someone whose colour theory clashes with central intuitions about colours would always face the task of showing how those intuitions can be explained (away), consistently with colours being the properties she claims they are. So it seems fair to require that any position should aim to accommodate, or at the very least explain, the intuitions as far as possible. (Those of us with internalist leanings, or just a hunch that intuitions are the best we have to go on, will generally prefer accommodating to explaining away.)

Intuitions about colours give no clear verdict on the location issue, but support both physicalist and dispositionalist (and to a lesser extent subjectivist) views. They can be divided into two broad classes, one highlighting the objective side of colours, and one highlighting their connection with human responses. The ‘objectivist’ class includes intuitions like the following: Colours are properties of objects. If you want to know the colour of something, you should look at the object, not in your mind, and try to bring yourself into a better viewing position, ask people with better colour vision than yours, etc. If the viewing conditions or your visual apparatus are less than optimal, your subjective colour experiences can come apart from the colours things really are (in such cases, the real colours follow the physical properties of the objects, which supports the claim that the colours should be located with physical properties). Indeed, colours seem to be independent of observing subjects to the extent that if humans became extinct, most things could remain the same colours as before. The most important intuition on the ‘objectivist’ side is perhaps the intuition that colours are causes – most importantly, causes of colour experiences. Intuitions about causal efficacy are employed in a central group of arguments for low-level locations which will be described later in this chapter.

The complementary class of intuitions – intuitions that highlight colour appearance – includes what might be called the transparency intuition (because of its similarity to intuitions about the transparency of one’s own mental states): If the object looks red in good conditions (normal daylight, appropriate distance and

orientation, normal visual apparatus etc.), there is just no further question whether it really is red. All it takes to have a certain colour is to look a certain way in good conditions. So if you want to know the colour of an object, you should not take it into the lab to investigate its microphysical properties, but simply look at it in good viewing conditions. Justified beliefs about the colours of objects are available simply on the basis of colour perception in good conditions; this, and not scientific investigation, is the highest court of appeal.

Another cluster of intuitions that highlight colour appearance are intuitions about the relations among colours in the 'colour solid', i.e. along the three axes of hue, saturation and brightness. One is the intuition of unity: if two samples, e.g. Zinka the canary and a good colour photo of her, look the same colour in normal (or most or all) lighting conditions, there is no further question whether they really are the same colour. The same is not true for their physical properties, which suggests that the latter are not the colours. Other intuitions in this group concern relations of similarity and difference, e.g. that orange is more similar to yellow than to blue. All these relations hold for colour appearance, but not for the corresponding physical qualities.⁷

These two stories about colours sound uncontroversial enough when considered in isolation. But they are incompatible with each other. For a start, if colours are the physical properties of objects causally responsible for colour experiences, there is no guarantee that they display the relations of similarity, difference and identity we expect of the colours. In fact, it turns out that they don't. Starting from another angle, if all it takes to be yellow is to look in a certain way in certain conditions, and if at the same time colours are to be physical properties, then we would need a guarantee that a physical common factor was in place in every instance. But no such guarantee is forthcoming. In other words, the properties that satisfy the intuitions of transparency and unity etc. are not the same as, and are not guaranteed to coincide with, the ones that satisfy the intuitions about causal efficacy.

⁷ The status of this group of intuitions as core intuitions knowable a priori by everyone competent with the colour concepts is disputable; perhaps they are just deep empirical truths. However, even if this is the case, they deserve to be taken seriously; a theory that rejects them owes an explanation of their plausibility and universal acceptance.

So there seems to be no perfect solution to the location problem about colours. We have two groups of intuitions pointing opposite ways, with no immediate prospects of reconciliation. The picture I've presented is simplified, of course, but filling in the missing details promises to make things worse, not better. The situation will become even more complicated when we take more candidate properties into account, e.g. dispositional physical properties (reflectance spectra) which are clearly relevant, but satisfy neither causal intuitions nor all the intuitions about similarity relations etc. What should we say about this messy situation?

1.2 A diagnosis

Our picture with several different levels of relevant features suggests a diagnosis of how this situation has come about: colour concepts equivocate between several different levels of properties which coincide in the conditions that count as favourable for colour assessment, and which, therefore, are easily conflated. In conditions favourable for colour assessment, a red object looks red, is disposed to look red, and has one of the surface structures and reflectance spectra characteristic of red objects. Indeed, it is natural to suppose that the reason why these particular conditions count as favourable is precisely that they are the conditions in which different levels of relevant features coincide with sufficient regularity. In other words, they are the conditions that secure that the characteristic high level properties (colour experiences and dispositions to elicit these) point us to the characteristic properties (sometimes multiple ones) on lower levels. In conditions favourable to colour determination, colour experiences are our means of tracking facts about surface structures – or rather, facts about differences in surface structures; this is why colours are such a handy tool for navigating in the world.

Because all the characteristic features coincide in C-conditions, there will be less of a motivation to distinguish clearly between the different levels for ordinary purposes. This might explain why we are not normally aware that there is a problem with contradictory requirements in our colour concepts. Location problems become pressing mainly when we do philosophy. Even if the different levels of properties associated with colours are not distinguished in everyday life, they are not to be conflated. In other possible worlds and outside favourable conditions, they might

come apart. For example, Zinka the canary will look red if observed in red light. And things could have been such that the (actual) surface structure of her feathers normally gave rise to colour experiences such as the ones we actually get from looking at the sky. So even if colour concepts equivocate between levels, there is an interesting location dispute to be had: If or when the relevant properties on different levels come apart, which ones should the colour concepts follow?

This question may not have a clear-cut answer. If indeed everyday colour concepts equivocate between levels, one might argue that we should simply leave the matter there, and that a location with one of the candidate properties would be a stipulative sharpening of our concepts rather than a faithful account of them (more on this strategy below).

1.3 Other examples of levels

Similar structures with different levels of relevant properties are discernible in many areas other than colours. A straightforward and frequently used example is that of heat, with phenomenal experiences of heat on the highest level, molecular mechanical energy at the lowest (or the lowest yet known), and presumably others in between, such as perhaps dispositions to elicit heat experiences, dispositions to affect mercury thermometers in certain ways, etc. Most people, at least among philosophers and scientists, seem to favour a location with physical properties on the level of molecular mechanical energy, and claim that heat just *is* molecular mechanical energy. One might argue that even if this is the primary meaning of heat, a phenomenal concept of heat can be, and is, in use beside it.

Another, notoriously difficult example are mental states. In this area, we have much the same structure as with colours, and all the same moves are available (only, the problems seem a lot harder with mental states, probably because the issue is so 'close to home'; we could embrace eliminativism or causal impotence as conclusions about colours, but the analogous conclusions about mental states are unpalatable). At one extreme, we have phenomenal states, or qualia – what it feels like to be in a

certain state.⁸ At another, there are low-level physical states of brains. Between these, several other relevant levels can be distinguished: higher level physical properties, biological/ physiological properties, functional properties etc.

Again, intuitions (/folk-psychology) and the relevant scientific knowledge favour more than one location. Mental states seem to be states of brains because they are intuitively causally efficacious, and the states doing the causal work are arguably the physical ones. Also, mental states can be manipulated by manipulating the brain and the rest of the body, e.g. by ingestion of pain-killers, anti-depressants, or hallucinogenous drugs, or by exercise, relaxation techniques etc.

But mental states also seem to be functional properties; what makes a certain state a belief that it will rain this afternoon is presumably its relations to behaviour (e.g. wearing a waterproof jacket) and other mental states (such as concern about the hike planned for tomorrow and the fresh paint on the garden shed).

And, finally, (many⁹) mental states seem to be phenomenal properties because the essential thing about a pain or an experience of red is that it feels in a certain way. If something feels the right way, then it is a pain, or an experience of red, no matter what the physical story is, and if different physical properties had been the ones to accompany the experiences, then these would have been the ones relevant to pains or colours.

At least *prima facie*, the diagnosis given for colours seems attractive for mental states as well. Several levels of properties seem relevant to our ordinary concepts of mental states. As with colours, the levels are not clearly distinguished for everyday purposes. All locations have intuitions pointing towards them, but none of the levels can satisfy all intuitive demands if taken in isolation.

⁸ Some, such as Dennett (1988) and Tye (e.g. 2002), deny the existence of qualia (at least if these are construed in a certain way). But even if (implausibly in my view) qualia should ultimately be eliminated, nothing seems lost by including them in a preliminary list of *prima facie* relevant properties. If you don't believe in qualia, you can pretend that that's what I am doing here, and this section should make sense.

⁹ This goes for a large group of mental states, but as is familiar from Wittgenstein's discussion in *Philosophical Investigations*, it does not hold for intentional states like believing, understanding, etc.

These examples should suffice to convey an intuitive idea of levels. The following sections aim to give a brief overview of various location options, some of the standard arguments for them, and some of the problems they face. This overview should also make the notion of levels clearer by relating it to a range of familiar arguments, and showing a bit of its potential in clarifying our picture of the realism disputes in question.

The paradigm of location disputes and of different levels of relevant properties may not work for all realism disputes. For example, it is hard to see how it would work for domains like abstract objects (numbers, mathematical facts etc.) or modality. It is mainly relevant where there are supervenience relations to a class of physical facts, but where it is not clear that these facts are identical with the disputed phenomenon. But where it does work, it promises to provide interesting results.

2. Location options

In response to the location question for a given domain, several strategies are available. The disputed phenomenon may be located on a low or basic physical level, or on some higher physical level, or with functional properties, or on the level of response-dispositions, or on the level of phenomenal experiences, or with some more complex choice of features of these or similar kinds.

Beside these general location strategies, there are various approaches to location disputes that do not consist in a straightforward choice of location. One might argue, with eliminativists, that the concepts fail of reference altogether. Or one might argue, with primitivists, that the phenomenon should be regarded as primitive and cannot be located with, or reduced to, anything else.

Also, one might argue that the location question has no clear-cut answer. One such view – we can call it the ‘equivocation view’ – would be that we should recognise a level equivocation in the disputed concepts and leave the matter there, rather than trying to purify the concepts of the equivocation. Another option is a multi-level location of some sort, e.g. by way of a functionalist story involving both role and realiser properties, and thereby accommodating more than one level of relevant features.

In the following sections, I will provide an overview of location options and their advantages and drawbacks. I will be speaking in very general terms, and many of the options outlined will make sense for some domains but not for others. The aim is just to convey a sense of the landscape; in order to do more than this, we would need to look at the detail of the arguments in particular domains. The various location options will be treated under two general headings: higher level locations and lower level locations. This division is merely for the purpose of quick and easy exposition, as the location options in each group generally share a lot of standard arguments for and against them. I do not mean to suggest that the spectrum of location options can be reduced to two polar options; it is, indeed, a spectrum.

2.1 Lower level locations (high, low, or basic physical levels)

We shall start with locations with physical properties. To make it a bit clearer what is meant by a *physical* property (or level), we should distinguish between two different senses of physicalism. First, there is a sense in which physicalism for a given domain entails that the disputed concepts refer to physical properties, and that the disputed properties are identical with physical properties – that colours, for example, are surface structures, or perhaps reflectance spectra. Secondly, there is a broader sense of physicalism which entails only that the disputed phenomenon is realised by, or constituted by, or supervenient on, or determined by, the physical (these hard-to-define relations will be discussed briefly in Ch. 3). The latter kind of physicalism is most interesting as a *global* thesis, often formulated along the lines of ‘if the physical facts about our world are determined, all the facts about our world are determined’ (this is a paraphrase of Chalmers 1996, p. 42), or ‘a minimal physical duplicate of our world is a duplicate simpliciter of our world’ (Jackson 1994, p. 164). By contrast, the first brand of physicalism seems most interesting in local versions, since a global claim that all terms that refer at all refer to physical properties seems highly unlikely to be true.

When I talk about low level locations in the following, I am talking about physicalism in the first sense: identification with, or location of the reference with, physical properties. Rejection of physicalism in this sense is compatible with physicalism in the second sense. One can thus be a (global) materialist without going

for a low level location in each and every location dispute, as long as one holds that all high level phenomena are physically realised, or supervene on the physical in the right way (provided that these relations are understood in a way that does not itself entail a low level location).

A first thing to notice about physical level locations is that there are many of them; for colours, for example, both surface structures and reflectance spectra are perfectly respectable physical properties. (There are even more levels to choose from if we include the properties typically treated by the other natural sciences – chemical properties, biological properties etc. – presumably these are equally natural, scientific, non-anthropocentric etc. properties, though they do not belong to the domain of physics.) Thus, a physical level location does not always mean a basic level location. This allows us to distinguish two different sorts of physical level locations: basic level locations and non-basic level ones. The former category might be empty if the idea of a basic level is illusory (more on this in Ch. 3).

A standard argument for low or basic level location is *ontological economy*. By reducing as many seeming higher-level phenomena as possible to lower-level ones, we get fewer ontological categories, and a simpler and more unified picture of the world. Relatedly, the history of science is a history of successful reductions from higher to lower levels; what we think of as paradigmatic epistemic achievements consist in bringing this sort of order into our world view.

(Higher level theorists might dispute the relevance of these considerations by arguing that all that is needed for explanatory reductions of the relevant kind is (the right sort of) supervenience, while low level locations in the form of identity claims are not necessary. There is presumably a long and complex discussion to be had over these matters, but this thesis is not the place to have it.)

Another prominent group of arguments for low level locations is what could be labelled causal arguments. Many such arguments are variations over *the causal exclusion argument*.¹⁰ In the original version about mental states, the causal

¹⁰ Kim (1998) and various other works. The statement of the argument to follow is inspired by Crane (2001).

exclusion argument shows that there is an inconsistency across the following individually plausible claims:

1. Mental phenomena have effects in the physical world
2. All physical effects have physical causes which are sufficient to bring them about (i.e. the physical domain is causally closed)
3. Mental and physical causes do not over-determine their physical effects
4. The mental and the physical are distinct (this covers dualism in any form – everything that is not an identity theory)

Less formally, we want to say that mental phenomena do some causal work, physical phenomena do all the causal work, the two are distinct, and there is no over-determination. But we can't have all of these claims. In the classic formulations of the argument, the invited conclusion is that mental and physical phenomena are not distinct after all.

This argument can be carried over to other areas, e.g. colours and dispositions. Versions of the argument could presumably be stated for any area in which we have multiple levels, of which the lower ones are levels of physical properties, and where the higher level properties are thought to cause physical effects. Thus, it would presumably apply to various kinds of special science causation, e.g. causal claims within biology, chemistry, and physiology.¹¹

Note that if causal exclusion arguments do indeed apply to scientifically kosher higher level properties, e.g. biological, chemical, and higher level physical ones, the argument will count in favour of basic level locations only, whereas it will be no use as an argument for higher level locations within the physical domain. (So if there is no basic level – or if we have no independent evidence to believe in one, and refuse to take an a priori argument like the causal exclusion argument as a sign that it exists – then so much the worse for the causal exclusion argument.) Variations on the theme of causal arguments are Frank Jackson's argument for colour physicalism and Mark Johnston's missing explanation argument.¹²

¹¹ See e.g. Block (2003).

¹² For Jackson's argument against colour physicalism, see Jackson (1998), p. 90-95. For Johnston's missing explanation argument, see Johnston (1991) and Ch. 4.

2.2 Higher level locations (response-dependence, functionalism, phenomenal states)

Higher level locations are locations with properties that are not physical in the narrow sense, and which are often closely related to a human perspective. They include response-dependence theses of the sorts that concern not just the concepts but the subject matters under discussion, including Johnston's and Wright's, but not Pettit's versions. Location with functional properties, i.e. those versions of functionalism that locate the reference with role properties rather than their realisers, are also high level locations. Realiser functionalism, by contrast, is better placed under physical level locations, though it has some advantages over other views of the kind (more on this below). High level locations also include location with phenomenal properties, though this option, at least in its pure form, might be relevant only in the case of mental states. Finally, projectivist positions might perhaps be placed in this category. However, they (or at least some of them, e.g. Boghossian's and Velleman's projectivist view of colour¹³) differ slightly from genuine high level locations in distinguishing two questions that other positions answer together:

1) Where should the reference of the concept be located?

and

2) What level of properties is most essential to the phenomenon in question?

According to such projectivist/error theory combinations, the concepts purport to refer to properties on a relatively low level which satisfy all intuitions connected with the concepts. But as no such properties exist, the concepts fail of reference. In reality, the feature most essential to colours is colour experiences, and contrary to what we might have thought, these are all there is to colours.

High level locations can be more or less subjectivist, depending on e.g. whether appearances can always be taken at face value, or whether there are some constraints (C-conditions) that they must meet in order for the connection between appearances and facts to be secured. Most high level locations, if chosen globally, will entail a strong form of subjectivism (with the possible exception of functionalism).

¹³ Boghossian & Velleman (1989).

Standard arguments for high level locations include transparency arguments, arguments from unity, and various (other) conceivability arguments, e.g. the zombie argument. However, anything said on this level of generality should be taken with a grain of salt; the arguments needed and used to support a given high level location will depend heavily on the subject matter and on the nature of the proposed high level theory. Most of the standard arguments have been developed for philosophy of mind, but many can be adapted to, and are familiar from, other domains.

The paradigm transparency argument is Kripke's argument against identity theories of the mental.¹⁴ The argument can be summed up as follows:

1. If A and B are rigid designators, and A is identical to B, then necessarily A=B
2. 'Pain' is a rigid designator
3. 'C fibre firing' (CFF) is (or is a placeholder for) a rigid designator
4. So if pain = CFF, then necessarily pain = CFF
4. If pain = CFF, then necessarily pain = CFF
5. Intuitively, there could be CFF without pain and vice versa
6. The usual strategy for explaining away the appearance of contingency in a posteriori necessities ('pain is CFF, but an epistemic counterpart of pain is not CFF') is not available in this case, because if something is an epistemic counterpart of pain, i.e. feels painful, it is a pain
7. So pain is not CFF

Or, in a little more detail: Identity theories of the mental make claims like 'pain is identical with C-fibre firing'. An identity statement involving two rigid designators has to be necessary. (This is why: a rigid designator refers to the same thing in all worlds (or all worlds in which they refer to anything, depending on your view on rigid designation); so if rigid designators A and B refer to the same thing, they will do so in every possible world.) 'Pain' is a rigid designator of pain; Kripke's justification for this claim is that if something is a pain, then being a pain is an essential property of it. A rigid designator can also be found for the corresponding

¹⁴ Kripke (1972), p. 144-155.

physical property; let CFF be a place-holder for the name of the physical property until we discover what exactly it is. So if pain and CFF are identical, they are necessarily identical. However, it is conceivable that there could be pain without CFF (or any physical property it might be replaced with), and vice versa. Furthermore, the usual strategy for explaining away the appearance of contingency in a posteriori necessities is not available in this case. The explanation would be something like this: 'pain is CFF, but an epistemic counterpart of pain is not CFF, and that's why pain without CFF seems conceivable'. But if something is an epistemic counterpart of pain, i.e. feels painful, it is a pain; this is the transparency intuition. So this won't do. The invited conclusion is that pain is not identical with CFF.

Analogous arguments can be constructed for other domains where the required materials are available: intuitions about transparency, two rigid designators A and B for the properties that are to be shown to belong on different levels, and intuitions that these levels can come apart. Also, the argument could presumably be applied, not only against physical level locations, but also against e.g. location of mental properties with functional properties as opposed to phenomenal ones (let CFF stand for a functional property and see what happens). Transparency arguments are very important for certain kinds of response-dependence theses.

Another group of standard arguments for high level locations are what we could term arguments from unity. The overall structure is this: Intuitively, being F (red, painful, etc.) is one thing, not several. But on the level of physical realisers, there is multitude rather than unity – either across actual cases or across possible worlds. So this level can't be the right place to locate the phenomenon. The multiple realisation argument against identity theories of mental states is an argument of this sort: 'pains can't be identical with CFF because pains could be realised by different physical properties'. So is the argument that 'red can't be identical with surface structure S1, for other red objects have surface structure S2, S3 etc.'. Similar arguments can be applied against other locations than physical ones. For example, Putnam applies a version of the multiple realisation argument against the

functionalist view of mental states that the argument was originally designed to defend.¹⁵

I've advertised conceivability arguments such as the Zombie argument¹⁶ as a third group beside transparency and unity arguments. This label may be misleading, since conceivability premises play a crucial role in (versions of) both of the latter types of arguments. However, there is a certain kind of conceivability arguments that purport to show more than just non-identity between higher level properties and their physical correlates; these are arguments that the higher level properties fail of (logical) supervenience on the physical. An example is the zombie argument employed by Chalmers and others: It is conceivable that there could be beings with just the same physical and functional characteristics as you and me, but without any accompanying phenomenal states. So the physical and functional facts do not determine all the facts about consciousness; phenomenal states fail of logical supervenience.¹⁷

Of course, this brief survey is by no means exhaustive. For example, Jackson's knowledge argument against physicalism about the mental hasn't been mentioned, and doesn't comfortably fit into any of the described categories.¹⁸ I could go on – but this was supposed to be a brief introduction.

Speaking very generally (and at the risk of over-generalising), advantages of high level locations lie in their ability to accommodate intuitions about transparency,

¹⁵ For the multiple realisation argument against identity theories, see Putnam (1967). For the version against functionalism, see Putnam (1992). For a unity argument against colour physicalism, see Johnston (1992).

¹⁶ See Chalmers (1996) for a vivid statement of the argument (p. 94-99) and, later, an impetuous discussion of its implications.

¹⁷ This argument is more controversial than the other two. First, some think that logical supervenience is not what should be at stake. Secondly, many question whether we can really conceive of zombies, or whether our so conceiving has any relevance for the sort of possibility we should really be looking for.

¹⁸ See Jackson (1986). In brief, the argument is this: Mary, a scientist with unlimited cognitive powers and knowledge of all the physical facts about colour, has been brought up in an environment where everything is black, white or grey, and has never seen other colours. When she sees those colours for the first time, she learns something new. So all the facts about colours are not entailed by the physical facts about them; colours do not logically supervene on the physical.

unity, and the possibility that the disputed phenomena and their realisers can come apart across possible worlds in certain ways. The problems about high level locations are with accommodating intuitions about causal efficacy, and perhaps with ontological economy. Of course, the details will depend on the subject matter under consideration.

2.3 How to choose a location?

How to choose the right location of a disputed concept? Two groups of desiderata have already been mentioned: scientific facts about the candidate properties and intuitions about the disputed phenomenon. Other things equal, intuitions of unity and transparency point to a high level location, while intuitions about causal efficacy point to a low level location (at least if the classic arguments from the literature get things right).

A third place to look for hints about location (possibly a subclass of intuitions) is at the *methods of detection of the phenomenon*. First, we should consider what criteria are most important for pragmatic purposes. If measurement of wavelengths became our primary basis for making colour judgements in everyday life, this would presumably move the reference of colour concepts from the level of response-dispositions (if that's where it is) to the level of reflectance spectra. (In that case, we could say that the reference changes, or that it has always belonged on the deeper level which is now duly recognised; both could make sense, depending on the circumstances.)

Not only pragmatically important criteria, but also 'ideologically' important ones matter to the location issue. The criterion considered as the most accurate way of detecting the phenomenon – *the highest court of appeal*, so to speak – would be a good guide to location. (An example: the highest court of appeal for colour determination is not measurements of wavelengths, but colour experience in C-conditions. And the higher accuracy of wavelength measurements can do nothing to remove the vagueness of colour terms. This suggests a high level location for colours.)

As mentioned in the 'diagnosis' above, a very direct way of asking and trying to answer the location question is this: when the levels come apart, which one of

them does the disputed phenomenon intuitively follow? In order to get a comprehensive picture, we should take into account as many possible scenarios as we can. With colours, for example, the picture is mixed: when physical properties and colour experiences correlated in C-conditions come apart in the actual world, the explanation is usually deviant lighting conditions, deficiencies in the observer etc., and the ‘real’ colours intuitively follow the physical properties. But when we look at the matter across possible worlds, the picture is different: If we consider a situation where colour experiences are correlated with different physical properties, the colours intuitively follow the colour experiences. The natural way to describe the situation is that colours have different physical correlates, not that colours fail to cause the usual kind of experiences, and these experiences are caused by something different from colours.¹⁹ So it is not enough to look at one kind of case where the relevant features come apart; to answer the location question, we should take into account as many cases and corresponding intuitions as possible. The same goes for arguments for location options; none of them should be treated as knock-down arguments in their own right. What matters is the total picture of arguments for and against the various location options and (non-location) alternatives. Only by taking all these parameters into consideration can the location issue be properly addressed.

3. Alternatives to location

3.1 Error theories: aiming too high

Beside the location options discussed, there are various approaches to location disputes that do not consist in choosing a location. One such approach is error theories: concluding that the disputed concepts fail of reference. Error theories might

¹⁹ Two-dimensionalism might offer colour physicalists a way of explaining away the problematic intuition. The thought would be that if the considered world had been *actual*, different physical properties would have been the colours: the ones that elicit colour experiences in that world. This might make it seem that the colours follow the response-pattern across worlds, not the physical properties. But this is only because we think of the counterfactual world as if it had been actual; if we consider it merely as a counterfactual world, the correct description is that in that world, colours do not cause colour experiences, but that colours are the same physical properties in that world as they are in the actual world. Analogous arguments could be made in other domains.

be motivated in (at least) two ways. One is by a recognition that intuitions about the subject matter under discussion point to different levels, and that no one level satisfies a sufficient proportion of the intuitions to deserve the reference of the concepts. For example, one might hold that colour concepts support intuitions about unity and transparency that, together with the scientific facts about colours, would place the colours on a high level, and that they also support causal intuitions and intuitions about colours being properties of objects that would place them on a low level. As no single level of properties satisfies all these intuitions, the colour terms fail of reference, and colour ascriptions should be explained as results of a projective error (as in Boghossian & Velleman (1989)). Alternatively, an error theory might be motivated by the consideration that a certain concept definitely aims to refer to a certain level and would become useless if it turned out that no appropriate referent was to be found on that level, combined with the empirical result that no such referent exists. An example of the latter kind is *phlogiston*.

Error theories provide a nice, clean way of dealing with complex problems. In some cases, such as that of phlogiston, they are certainly appropriate. However, in cases motivated in the first of the ways described above, one might question the strategy; if no perfect referent is found, why not locate the reference with the best approximation, or even conclude that the concept refers to a complex of different properties on different levels, rather than concluding that the concept fails of reference? One might feel that this kind of error theories demand too much of concepts in order to grant them reference.

3.2 Primitivism: stipulating a level that has it all

Primitivism is another view that doesn't represent a straightforward choice of location within the set-up described. On this view, the disputed concepts refer to sui generis properties, not to be reduced to, or conflated with, properties of any other kind. Primitivist positions such as Campbell's (1993) on colours share with eliminativism the view that the referent of the disputed concepts must be a property that satisfies all the intuitive demands on the concept. They also agree that none of the levels of candidate properties we have considered meet these requirements. But one man's modus tollens is another's modus ponens; where the eliminativist

concludes that the concepts fail of reference because none of the candidate properties satisfies all intuitions, primitivists conclude that there must be other properties for the concepts refer to – a range of unique, simple properties different from the usual candidate properties and specially designed for the purpose, so to speak. In the levels terminology, one might say that this sort of primitivist position takes advantage of the fact that levels are not always nicely and unequivocally individuated, and conveniently invents a level that has everything we need (or maybe postulates a property that is outside the level framework, and which meets the requirements). The advantage, of course, is that all intuitions will by hypothesis be accommodated. The cost is a loss in terms of ontological economy, and worries about the strategy of stipulating what we need, regardless of how it fits with our general picture of the world. Also, there will be worries regarding ‘competition’ between the primitive properties and the traditional location candidates; for example, causal exclusion worries regarding the primitive properties and their physical competitors for causal work would need to be addressed.

Perhaps the term primitivism should be allowed to cover another sort of position as well: the view that the disputed properties, e.g. mental states, are a species of their own, and not reducible to, or logically supervenient on, physical states. Chalmers’ (1996) view of mental states would be an example, as would traditional Cartesian dualism. This version, too, would have a cost in terms of ontological economy, but would not have the objectionable flavour of stipulation (at least in Chalmers’ case). Worries about causal efficacy would still apply. Note that a view like Chalmers’ is really a location option, locating the disputed phenomenon with phenomenal properties. Thus, unlike the first version of primitivism described, it is not a way of side-stepping the location issue.

3.3 Multi-level locations: functionalism

It may be an over-simplification to think that location disputes enforce the choice of a single level as the correct location of the disputed phenomenon. This is a central idea behind what we could call multi-level location views: views that locate the disputed phenomenon not on one, but on two (or more) levels. Paradigmatic multi-level location views are functionalist views. A functionalist view would take a

disputed phenomenon to comprise two different levels: a role level with a functional property, and a level of physical realiser properties. For colours, for example, the relevant role property could be that of having a property that causes certain familiar experiences in C-conditions, makes the object appear more similar to some (yellow) objects than to other (blue) ones in a certain respect, etc. The realisers would be the properties that fulfil this job description – probably microphysical properties of the surfaces of objects. Such a combination of properties would arguably do a better job of accommodating intuitions pointing to different levels than any single level location. The realisers would account for the causal intuitions, and the role-description could give us transparency and whatever else is built into the ‘job description’ for the concept (in my example above, a version of unity).

A functionalist theory would have significant advantages over a simple physicalist location. First, it would provide an answer to the transparency challenge. Having a certain range of effects on appropriate observers would be part of the role description for ‘red’, so in order to count as red, a surface has to appear in the right way to standard observers in C-conditions. With the unity challenge, the functionalist would also fare better than simple physicalist locations. First, because the different realisers of redness, however different they may be, are at least united by the role property they have in common. Secondly, a version of intuitions such as those about the relations of identity, similarity, and difference in the colour solid could be built into the role description too. It would have to be a modified version, however, stating that red *appears* more similar to orange than to blue etc., and not that it *is* more similar...; the latter requirement would entail reference failure, since the actual realiser properties do not in fact display these relations.²⁰ So a functionalist view would be able to accommodate many, though not all, of the intuitions about colours. Similar advantages might be expected in other areas.

The problems with functionalist theories include various domain-specific arguments, most notably about mental states.²¹ But because these problems are domain-specific, we won’t go into them. There are also some more general worries. A first challenge is this: why can functionalists get away with locating a phenomenon

²⁰ See Hardin (1988) p. 2-7 for a summary of the relevant scientific facts.

²¹ See Chalmers (1996).

with two levels of properties if others must choose? After all, roles and realisers are different properties – properties that can come apart. So it is not clear that we can allow e.g. redness to be both at once. If a choice is enforced, many advantages of functionalism will be undermined. A location (solely) with role properties will raise most of the problems typical of high level locations, while a location with realiser properties will raise (at least some of) the usual problems with low level locations. If, on the other hand, we allow a two-level location in the case of functionalist views, why shouldn't we also allow other views to claim the benefit of taking into account both a high, relatively anthropocentric level and a level of physical realisers, saying e.g. that redness is the combination of a phenomenal property and its physical correlate? In response to this challenge, the functionalist might reply that she offers an account of the relationship between the two levels she recognises: the higher level property is the property of having the causal role played by the realisers. By contrast, if the suggestion is just to combine two (or more) different level properties, there may not be a story to tell about how they are connected, or how to find the appropriate 'realisers', or even how physical properties could act as 'realisers' for the higher level properties in question. We shall touch on these issues in chapter 3.

A second, serious challenge to functionalism arises from the consideration that it reserves a privileged position for a basic level of realisers. Unlike the higher level properties of other types of accounts, the 'higher order' role properties of functionalism are not supposed to be a separate level of properties in their own right; they are simply properties of having certain other properties. The properties doing the real work are the realiser properties. Even though some roles might be realised by other functional role properties realised by further realisers, the chain has to come to an end somewhere. The position works best with the presupposition of a level of ultimate realiser properties. If this presupposition is given up, it opens the possibility of an infinite hierarchy of properties of having properties of having properties..., and this picture is very unattractive (whereas an infinite hierarchy of levels might be acceptable on other conceptions of levels; see Ch. 3).

There are two problems with the idea of a basic level. First, even if there is a basic level of ultimate realisers, these will be given a privileged position, while other properties will be relatively unreal in comparison. This seems intuitively strange,

especially if the realisers are e.g. elementary particles that some future science might tell us about, but which we don't know from ordinary life. This would mean that all the properties we are acquainted with are really properties of having properties, or properties of having properties of having properties..., which sounds intuitively unsatisfactory. Secondly, and worse, the presupposition about a basic level is arguably unwarranted, and may be false, as we shall see in Ch. 3. And depending on empirical luck – that after all there will turn out to be a basic level – is not an attractive move. This unsettled issue about the basic level poses a serious problem for functionalism. Unfortunately, we can't do justice to these matters in this short introductory section. We shall return to some of the issues in later chapters.

3.4 Equivocation views

If something like the diagnosis suggested at the beginning of this chapter is right for a given domain – that features on several levels are relevant and coincide in C-conditions – one might feel that the project of choosing a location is misguided. Why not simply acknowledge the equivocation in the concept, and let the matter rest there? The conflicting intuitions are nicely explained by the thesis that the concepts equivocate between phenomena on different levels which are all relevant, correlated in C-conditions in the actual world, and potentially divergent outside C-conditions and in other possible worlds. One might think that the further step of singling out one level as the all-important one would achieve only a stipulative sharpening of the disputed concepts, rather than a faithful and illuminating account of the phenomenon under discussion. Let us label this sort of view the equivocation view. I shall discuss it in a bit more detail than I have done with the other views, as this kind of view is not often encountered in the literature, and hence will be less familiar than the others. An example (or near enough) is found in Chalmers (1996), who endorses something that sounds very much like an equivocation view about pain:

One can tie oneself into all kinds of knots by worrying about whether the phenomenal quality or the functional role is more essential to pain. [...] There is little point in trying to legislate matters one way or the other: nothing important rests on the semantic decision as to whether some phenomenal quality is really essential for something to count as pain. (1996, p. 17)

If an equivocation view is generally the correct stance to take, this would explain why many people – including many philosophers – fail to see the point in location disputes such as the one about colours. The disputes impress them as battles over words, with no substantial issues at stake and no interesting conclusions in view. The thought seems to be something like this: The disputed phenomenon has all these different aspects, and we know the facts about each of them. Why quibble over which of the aspects deserves to be the sole heir of the colour title?

The equivocation approach is also supported by the following considerations: The features that constitute arguments for the various general location options (physicalism, response-dependence theories, functionalism, projectivism etc.) are present to different degrees in different areas. For example, colours support intuitions about the importance of underlying physical properties, and also intuitions about transparency that point to a high level location. The transparency intuitions are weaker for colours than in the case of mental states; they are relative to a set of C-conditions, and they have strong competitors drawing in the opposite direction. With pains and other (phenomenal) mental states, again both types of intuitions are present, but this time the transparency intuitions are stronger and very hard to give up. With intentional states (beliefs, desires, hopes, intentions etc.), transparency intuitions are still present, but they are a bit weaker than with phenomenal states, as there are examples that transparency sometimes fails; you can be wrong about what you believe, desire, hope etc. Summa summarum: different domains display combinations of very similar intuitions, but with different weighting.

Suppose we have two classes of disputed concepts, A and B. In both cases, there are intuitions pointing both ways, but for A-concepts the balance tips slightly in favour of response-dependence, whereas in case B the balance tips slightly in favour of physicalism. If we want single-level locations, we have to conclude that A-concepts are response-dependent, whereas B-concepts are concepts of physical properties – i.e. that A and B are very different kinds of concepts, describing very different kinds of phenomena. But this conclusion is counterintuitive given that intuitions about A and B are roughly alike, and differ only in the weighting of certain intuitions. An equivocation view, by contrast, would allow us to take seriously all the

components pointing to different levels, and not only the ones that point to the ‘winner’ of the location dispute. This is a strong advantage of equivocation views.

A possible disadvantage might be that seemingly simple questions like ‘what are colours?’ will have no simple answer.²² In response to this challenge, the equivocation theorist could say that there is an everyday sense in which this is a perfectly simple question with a simple answer: ‘red is that colour, blue is that one, colours are the features salient in those examples’ etc. But when asked as a philosophical question in the context of location disputes, it is less simple, and there is no reason to assume that it must have a simple answer. Indeed, the difficulties philosophers have in settling location questions suggest that it does not. The best answer in this context may well be a story about the various aspects relevant to colours, the way they work together, and the way we tend to equivocate between them. A recurrent theme in this work is that many ‘simple’ questions, inviting too simple answers, create more trouble than we would get by allowing a bit more complexity. Asking too simple questions can be a way of begging questions and creating pseudo-problems.

A second objection to the equivocation view might be that it charges everyday colour discourse with massive error. We think we know what property we refer to when using the concept ‘red’, that it’s a relatively simple property, etc.

One response to this challenge would be to deny the premise, and argue that we make no such presuppositions. A second defence would be that if an equivocation view contains an element of error theory, it is no worse off than its competitors. Most plausible views entail that some of our intuitions about colours get things wrong. Those who choose a single-level location must either explain away the intuitions pointing to alternative locations or assert (implausibly) that they don’t exist. The advantage of the equivocation theory is precisely that no intuitions need to be rejected. So in terms of error attribution, it does better, not worse, than the competitors. A third defence would be to argue that the view that concepts need to have a single-level referent in order to be in good standing is out of step with the way most of our concepts work. In order to do their job, colour concepts need to

²² This point was raised by Nikolaj Jang Pedersen in discussion. It would also apply to functionalist views.

equivocate, as the importance of colour judgements rests both with the lower level properties tracked by colour experiences in C-conditions and the fact that we can access them via these experiences. Moreover, as the ‘paradoxes’ of material constitution show, similar equivocations can be found in concepts whose status is not nearly as controversial as that of the colour concepts, e.g. the concepts *ship* and *cat* (see the appendix to Ch. 3).

A third objection to the equivocation approach might be that it has a quietist flavour and threatens to make a lot of seemingly interesting philosophical discussions irrelevant. If the realism dispute about e.g. colours is to be resolved by recognising that our colour concepts equivocate between a phenomenal, a response-dispositional, and one or more physical levels, then all the familiar arguments in favour of one location rather than another would seem to lose some of their interest, as would the theories they are arguments for.

However, it seems possible to retain the insights of an equivocation view while avoiding quietist consequences. The appealing feature of the view is that it allows recognition of several levels of properties and their contribution to the phenomenon in question. This idea is compatible with an interest in location disputes, since the latter can still tell us interesting things about the relative importance of the different components in the concepts. The best view, it seems to me, is a hybrid view that recognises and distinguishes the different components, and takes location arguments seriously as clues to the various levels in play and the weighting between them.

A cluster of views somewhat similar in spirit to the equivocation view would construe the ‘equivocal’ concepts as confluences of different single-level concepts. For example, the concept of heat might be construed as an uneasy mixture between a scientific concept of molecular mechanical energy with its characteristic scientific measuring methods etc., and an everyday concept for which the heat sensations installed in us to keep human bodies from harm constitute the primary mode of access. The idea of dividing the concepts into single-level components could be filled out in several ways. One view would have it that we should distinguish the single-level component concepts for philosophical purposes only, while the real-life concepts equivocate or mix levels, and should be allowed to continue to do so. A more revisionary view would be that new single-level concepts should replace their

confused multi-level ancestors. Another, non-revisionary view would be that everyday concepts do *not* equivocate, but that there are really two or more different concepts in use under the same name (with the context determining which one is relevant in each case), and that the conflation arises only when we do philosophy about the concepts and fail to recognise the exact way the concepts work. The common ground between these views is that they all allow different levels into the picture, with no need to regard one of them as privileged or all-important (though presumably the listed views are compatible with considering one or more levels more important than the rest).

Which of these alternatives is the best choice? Probably no general answer is forthcoming. It may well be that for some concepts, the actual practice is better described as one with two or more parallel concepts that refer to different-level aspects of the same phenomenon, while for others, a single, equivocal concept is in use. In the latter case, it might be useful to distinguish the component single-level concepts for philosophical purposes, or for other specific purposes, or quite generally, or not at all. All the options should be recognised. What (if any) concepts fit which options is a question for another occasion, though.

For what it's worth, I think most philosophically interesting concepts, and certainly the ones that give rise to heated location disputes, probably spread over more than one level. This will make some sort of equivocation view more attractive than a single-level location in many cases. This does not lead to quietism though; most of the arguments from location disputes make sense without the presupposition that one level must be singled out as the all-important one. They provide valuable clues to the various levels involved and their relative importance, and can thus provide a lot of insights, even if the aim is not always a single-level location. Furthermore, I think location disputes call for a 'holistic' approach: all intuitions and arguments should be considered, and it is the overall picture, rather than particular arguments, that should settle the disputes.

4. Summary

This chapter has presented a rough outline of the levels-based framework which will serve as backdrop for the discussion of response-dependence in later chapters. We

have seen how realism disputes can be construed as disputes over the location of the reference of the disputed concepts, where several candidate properties on different levels compete for the role as referents. We have revisited some of the standard arguments in favour of particular locations, and considered some alternatives to straightforward single-level locations – eliminativism, primitivism, equivocation views, and multiple level locations. Hopefully, this has conveyed some idea of levels and location; we shall need them later in this work. However, this first outline has left unexplored a number of central issues, many of which will have an impact on the use made of the framework later in this work. In the next chapter, we shall go into more detail with some of them.

Ch. 3: Some questions on levels

In ‘The non-reductivist’s troubles with mental causation’, Jaegwon Kim introduces the idea of levels in a way that raises several important questions. These questions set the agenda for this chapter. Kim writes:

The ontological picture that has dominated contemporary thinking on the mind-body problem is strikingly different from the Cartesian picture. The Cartesian model of a *bifurcated* world has been replaced by that of a *layered* world, a hierarchically stratified structure of ‘levels’ or ‘orders’ of entities and their characteristic properties. It is generally thought that there is a bottom level, one consisting of whatever microphysics is going to tell us are the most basic physical particles out of which all matter is composed (electrons, neutrons, quarks, or whatever). And these objects, whatever they are, are characterised by certain fundamental properties and relations (mass, spin, charm, or whatever). As we ascend to higher levels, we find structures that are made up of entities belonging to the lower levels, and, moreover, the entities at any given level are thought to be characterised by a set of properties distinctive to that level. Thus, at a certain level, we will find lumps of H₂O molecules, with such properties as transparency, power to dissolve sugar and salt, their characteristic density and viscosity, and so on. At still higher levels we will find cells and organisms with their ‘vital’ properties, and farther up organisms with consciousness and intentionality. Beyond them, there are social groups of organisms, and perhaps groups consisting of such groups. [...] Sometimes, one speaks in terms of ‘levels of description’, ‘levels of analysis’, or ‘levels of language’; the layered model is often implicit in such talk. (1993 s. 337.)¹

This passage raises a number of important issues regarding the levels framework. One is the question about the relation between levels: how are the relevant properties on one level related to those on the next level? Kim’s remark that higher level structures are ‘made up of’ entities belonging to the lower levels suggests that the relevant relation is mereological – a relation between parts and wholes. On the other hand, the remark about ‘a set of properties distinctive to that level’ might suggest something more than mereological differences. Other candidates are supervenience,

¹ 1993 s. 337.

realisation, constitution, and identity. I shall discuss this issue in section 1 below. The related issue of how to individuate levels will be treated in section 2. The claim about ‘levels of description’ etc. at the end of the quote raises another crucial question: Is the levels metaphor just that, a metaphor, a way of speaking (as the phrases ‘levels of description’ etc. suggest), or do the levels mirror real, ontological differences (as Kim’s examples of levels might suggest)? This question is the topic of section 3. The beginning of the quoted passage raises the question whether there is a fundamental level that enjoys a particular, privileged status in the framework. Kim’s contention that many people seem to think so seems correct; whether they are justified in doing so is another matter. Section 4 is devoted to this question. Section 5 concerns the question what constraints, if any, the levels framework imposes on the way we think about properties, and conversely. In section 6, I consider how the levels framework can be justified, and suggest some areas in which it might be useful.

Finally, Kim’s emphasis on ‘entities’ raises the question whether there are levels of objects, or just levels of properties. I shall be concerned mainly with levels of properties, which is all that is needed for the discussion of response-dependence. But in the appendix to this chapter, I shall say a little more about levels of objects, and how distinguishing different levels of objects might be helpful with respect to puzzles of material constitution.

1. What is the relation between levels?

What is the relationship between the phenomena (properties, objects, states of affairs) on one level and those on neighbouring levels? The answer to this question will depend on whether we are thinking about objects, properties or states of affairs, and on what sort of level differences we have in mind. As indicated above, one suggestion is that the relation between levels is mereological; lower level objects are simply parts of higher level objects. Different levels of properties can then be placed in the level ordering according to the entities they are properties of. Kim puts it very clearly in the passage immediately after the one quoted above:

Thus, the world as portrayed in the new picture consists of an array of levels, each level consisting of two components: a set of *entities* constituting the

domain of particulars for that level and a set of *properties* defined over this domain. What gives this array structure is the mereological relation of *being part of*: entities belonging to a given layer are mereologically composed of entities belonging to the lower levels, and this relation generates a hierarchical ordering of the levels. As earlier noted, this multi-layered picture usually carries the assumption that there is a bottom tier, a layer of entities that have no physically significant parts.²

There is certainly a sense in which molecules, atoms and sub-atomic particles occupy different levels. However, mereological relations are not the most relevant ones for our purposes. To return to the colour example, colour experiences, dispositions to elicit them, and colour-related physical properties are clearly not related as parts to wholes. The same goes for the other examples of levels from Ch. 2, including heat (phenomenal, functional, and physical heat) and mental states (phenomenal, functional and physical aspects).

There is a growing awareness in the literature that the issue of the relations between levels may be more complex than earlier assumed. Jonathan Schaffer (2003) distinguishes no less than four ways to think about levels. One is the mereological sense, which Schaffer describes as ‘the central connotation of the levels metaphor’.³ He distinguishes three further ‘peripheral connotations of ‘levels’’:
‘(b) a supervenience structure, ordered by asymmetric dependencies; (c) a realization structure, ordered by functional relations; and (d) a nomological structure, ordered by one-way bridge principles between families of lawfully interrelated properties.’⁴

It can be disputed whether these connotations are ‘peripheral’. But Schaffer’s list of relevant relations is helpful. It should be supplemented with a few more candidates.

² 1993 s. 337; his italics. The mereological conception of levels is also found in Oppenheim and Putnam (1958), a locus classicus for the idea of levels.

³ Schaffer 2003, p. 500. See also Hütteman & Papineau (2005), who distinguish between two senses of levels – roughly, mereological and non-mereological. Unlike Schaffer, they reserve the term ‘levels’ for the non-mereological sense, for which supervenience and identity are the candidate relations considered.

⁴ Schaffer 2003, p. 500. I shall not discuss (d) further, as I take it to be a variation of a supervenience claim.

One is constitution; it seems natural to think of e.g. a statue and the clay it is made of as belonging on different levels in a sense close enough to the others to be included in our discussion. Another is identity, as we should take into account the relation (or cluster of relations) that disbelievers in ontologically different levels might take to obtain between different level properties and objects (if there is such a thing as contingent identity, it may well be relevant here). (Another option for level sceptics would be eliminativism regarding the higher levels, but then it hardly makes sense to talk about a relation between levels at all.)

1.1 The non-mereological candidate relations

In the following, I shall risk a brief account of the non-mereological relations and the differences between them, though it must necessarily be superficial, and may be controversial. The task is complicated by the fact that terms like realisation, constitution, and possibly even supervenience, do not refer to well understood phenomena, but tend to function as place-holders for relations that are ubiquitous yet not well understood – identity-like relations that are different from identity.

Supervenience, realisation, and constitution have one thing in common that sets them apart from the part/whole-relation. In order to determine what things stand in the part/whole relation to each other, we need only look at the *actual* facts in a certain spatial region. By contrast, in order to determine whether two phenomena supervene on, realise, or constitute each other, we must take into account how things are in other possible worlds and/or at different times and places in the actual world; variation across worlds and situations is a tell-tale sign that the relevant relation is not identity, but one of the ‘identity-like’ relations of realisation, constitution or supervenience.

Supervenience, realisation and constitution differ from each other in that constitution is a relation between *objects*, while realisation is normally thought of as a relation between *properties*, and supervenience relations can hold between properties or states of affairs.

Constitution, as traditionally understood, denotes a relation between objects that occupy the same spatio-temporal region, but differ in their modal or past or future properties. Classical examples are a statue and the clay it is made of, or a cat

and its body, or a ship and the sum of materials it is built of. For practical purposes, the relata of the constitution relation count as the same object, but due to their difference in modal properties, they cannot be identical (at least not in the ordinary, strict sense, though some authors allow for a less strict notion of contingent identity⁵).

The notion of *realisation* is part of a functionalist two-level story involving role properties and realiser properties. A role property is normally understood to be the (second order) property of having a (first order) property or properties playing a certain, often highly complex, causal role. Realisers are the properties that occupy the roles and do the causal work in question.

Thus, the core claim of functionalism involves two ‘levels’, or ‘orders’, of very different properties.

The picture might perhaps be extended to include further levels – either by regarding some roles as realised by properties that are themselves realised by further properties, or by having several levels of role properties that are all realised by a basic level of realisers (though this version would require a story about why it makes sense to think of the role properties as arranged in a hierarchy of levels).⁶

Presumably the functionalist apparatus can be extended to cover roles that are not defined in purely causal terms, though this would change the spirit of the proposal as described above, and would require an account of how role and realiser properties are related if not by realisers filling causal role-descriptions. Also, it might be extended to include non-physical realiser properties (raising the same problems).⁷

⁵ See e.g. Yablo (1987).

⁶ However, the extension to further levels sits uneasily with core thoughts behind functionalism. The traditional functionalist story is an attempt to tie higher level properties to more physicalistically kosher properties, and properties the appeal to which can save intuitions about causal efficacy in the face of causal exclusion arguments. On this approach, higher and lower level properties are very different kinds of properties. Hence, to reiterate the story for more levels than two would presumably not preserve the original motivations (especially not if the assumption that there is a basic level must be relinquished; see section 4).

⁷ See e.g. Kim 1998, p. 83.

Supervenience is probably the relation between levels that is most central to this work. In the following, whenever I speak of levels and it is not clear from the context what sense I am talking about, it is safe to assume that it is the supervenience sense. One reason that supervenience is so well suited to capture the relation between levels in the sense most interesting for present purposes is that it comes in a wide variety of versions, sufficiently similar to say that we're talking about levels in a roughly unified sense, and yet sufficiently different to allow for the diversity of the issues under consideration. The core idea is that a group of (high level) properties, let's call them the B-properties, supervene on another group of (low level) properties, the A-properties, iff no two possible situations are identical with respect to their A-properties while differing in their B-properties.⁸ Supervenience claims come in local and global forms, depending on how broadly the 'situations' in question are understood. If the claim is that the A-properties of *individuals* (in a sufficiently broad sense of the term that includes objects, people, societies etc.) determine the B-properties of those individuals, we get *local* supervenience. If the supervenience claim concerns the distribution of A- and B-properties in *entire worlds*, we get *global* supervenience. Local supervenience entails global supervenience, but not the other way around (intuitively, if the B-properties of each individual could not be different without a difference in the A-properties of that individual, then the B-properties of all the individuals taken together cannot be different without a difference in the A-properties of all the individuals). On another axis of variation, the 'possible' in the supervenience schema can be filled out in several ways, depending on what worlds are considered relevant. It can be logical possibility, in which case we get logical supervenience. Or it can mean nomic or natural possibility, yielding nomic or natural supervenience – supervenience in all worlds where the laws of nature are the same as in the actual world. Thirdly, it can mean metaphysical possibility, yielding metaphysical supervenience.⁹ The distinction between logical, natural and metaphysical supervenience is independent of the one between local and global

⁸ This formulation is based on Chalmers (1996) p. 33; there are other ways to do it, but for simplicity we shall stick with this version.

⁹ Some, e.g. Chalmers (1996), doubt that there is such a third category, while others take it for granted.

supervenience; the versions can be combined as you wish. Besides these most central varieties, there are others. One is ‘weak supervenience’ which says merely that in the actual world, a difference in B-properties always comes with a difference in A-properties (i.e. a local supervenience claim restricted to the actual world; the global version would make little sense). Another is Blackburn’s ‘ban on mixed possible worlds’, which claims local supervenience of B-facts on A-facts within each possible world, but not across worlds.¹⁰ It is possible to design supervenience claims to suit a wide variety of purposes.

Though it has received surprisingly little attention in the literature, there is also scope for variation regarding the source of the supervenience – e.g. whether it is underwritten by requirements on how the *concepts* function (as in ethics), or by relations that obtain between the *referents* of the concepts.

Which versions of supervenience are relevant to which domains? Are e.g. biological or economical facts logically supervenient on physical ones, or only naturally or metaphysically supervenient? In which sense, if any, do mental facts supervene on physical ones? These are hard questions and provoke a lot of disagreement. Hence, there is lots of work to be done in deciding which supervenience relations (if any) are appropriate with which level-differences. Of course, one is free to combine different supervenience claims for different domains. For example, one can hold (like Chalmers 1996) that biological properties supervene logically on physical properties, whereas phenomenal properties supervene only naturally. Which version of supervenience is relevant is thus not a matter to be decided once and for all.¹¹

A supervenience claim is not the final word on the relation between two kinds of properties or states of affairs. It says something about correlation patterns between

¹⁰ Blackburn (1984), Ch. 6.

¹¹ Something similar is true when supervenience is combined with response-dependence; different supervenience relations may go with (and possibly even distinguish or define) different versions of response-dependence, depending of the choice of supervenience bases and the class of worlds quantified over. Unfortunately, the relationship between response-dependence and supervenience will not be investigated (much) further in this work. To carry out such an investigation properly would require detailed investigation into particular domains, and this falls outside the scope of this investigation, though it is the natural next step to take on the basis of its results.

them, but it says nothing about why, or in virtue of what, the correlation pattern is as it is. In that sense, supervenience claims state problems rather than providing solutions. This, of course, is a valuable contribution in itself. But it means that, like ‘realisation’ or ‘constitution’, ‘supervenience’ is a sort of place-holder for the explanation we seek, but is not the explanation itself.

Finally, *identity* – the main rival of the idea of ontological level differences, and the relation that will be relevant if level differences are merely differences in mode of presentation.

In David Lewis’s terms, the identity relation is easy to explain: everything is identical with itself and not with anything else. The relation is more fully characterised by Leibniz’s law: If A is identical with B, A and B share all their properties. Some also endorse the principle of the identity of indiscernibles: If A and B share all their properties, then A is identical with B.

Some authors hold that identity can be relativised to times, or to concepts /sortals, or to worlds. This means restricting the range of properties that A and B must share if they are to count as identical. For *tensed* identity, A and B may be identical at t even if they differ in their properties at other times. For *contingent* identity, A and B may be identical even if they differ in other possible worlds. For *sortal relativity*, the idea is that A, say a statue, and B, say the lump of material it is made of, can be the same material object but not the same work of art (i.e. share all properties relevant to the first sortal but not all those relevant to the second).

With respect to the levels framework, the distinction between contingent identity and strict identity is especially relevant. The traditional view is that if A and B are identical, they are so necessarily. The reasoning is as follows: Anything is necessarily identical with itself. So, in particular, necessarily (A=A). If A is identical to B, by Leibniz’ law, they share all their properties. One of A’s properties is this: necessarily (_ = A). So B, too, must have this property, and hence be necessarily identical to A. But if this is the case, then e.g. statues and the lumps of material they are made of can’t be identical, since they can come apart in other possible worlds. They might, instead be viewed as objects at different levels, related by the constitution relation. Or one could argue that contingent identity is possible, and relevant to such cases. Contingent identity might be compatible with level

differences in a certain, fairly weak sense. But identity in the strict sense is not a relation between different levels; it is the negation of the claim that there are level differences.

1.2 Different cases, different relations

These relations are not competitors for a single job as *the* relation between levels; some may be relevant to some cases, and others to others. Some of the relations – especially supervenience and realisation – can even coincide for a particular set of neighbouring level properties. If property A is a realiser of property B, this is consistent with B supervening on A; indeed, the supervenience claim might be expected given the realisation claim. Likewise, if object B is constituted by object A, states of affairs involving B will presumably supervene on those involving A.

Identity rules out the existence of levels in any ontologically interesting sense, and so presumably rules out realisation and constitution (depending on their exact formulation). For supervenience, the picture is mixed: identity entails certain symmetric supervenience relations (if A and B are identical, a difference in B will unproblematically entail a difference in A, and conversely), but rules out *asymmetric* supervenience claims.

For brevity, I will often be talking about ‘levels in the supervenience sense’, which is meant to include realisation, constitution, and supervenience relations other than those that obtain for identical entities. This – as distinct from the mereological variety – is the sense that will concern us in the discussion of response-dependence. Where I talk about levels without further specification, this is what I mean.

To complicate the picture further, concepts can equivocate between different levels of properties or objects (as described in Ch. 2), and *different relations between levels might be relevant for different disambiguations of a pair of neighbouring level terms*. For example, ‘molecule’ can arguably be understood in a ‘mereological’ or ‘physical’ low-level sense in which molecules have particular atoms as *parts*, and a ‘functional’ sense in which a being a molecule is a property with a functional essence, and the collection of atoms act as *constituents* rather than parts.

Likewise for ‘table’ (though in this case the functional connotations are somewhat stronger than in the molecule case); a table can be thought of as the sum of

its parts, but also as a thing with a functional essence which may be *constituted* by its parts, but is different from the sum of the parts.

As with many of the distinctions in this work, we're talking in-principle-options when distinguishing mereological, functional and other senses of the terms; real-life concepts of objects and properties may well equivocate between different options. Thus, the question of the relation between levels may not always have a clear answer.

Do levels related by these various relations comport, in the sense that if two properties or objects are on different (/adjacent) mereological levels, they are also on different (/adjacent) 'supervenient' ones? It seems reasonable to expect that they do, at least to a certain extent. One reason to believe this is that in many cases, the same pair of concepts will cover neighbouring levels in more than one sense, as in our examples above. However, it is no settled matter that the senses of levels always comport. Some (e.g. Kim 1998, p. 82) argue that they do not, whereas Schaffer thinks they do, and that if not, the idea of levels should be given up.

1.3 A unifying sense of levels?

If there are all these different senses of levels, how come people have been talking seemingly sensibly about levels for so long without distinguishing the different senses? And how could my claims about levels in Ch. 2 make sense without a specification of what sort of levels I was talking about? If levels in the different senses comport to a sufficient extent, this would provide one explanation of how we have so far got away with conflating them. Another possible explanation would be to hand if it turned out that there was an intuitive idea of levels that somehow united the various senses of levels – a sort of common factor that made it sensible to treat them all under the same heading.

What might such a common factor consist in? An obvious suggestion is an ordering of the levels according to their closeness to, or distance from, a subjective standpoint. For the colour example, there is a clear sense in which we move further away from the subject in the ordering: 1) colour experiences; 2) dispositions to elicit these; 3) reflectance spectra; 4) micro-physical surface structures of objects. This is true in the sense that 1) is a property of subjects, while 2) to 4) are properties of

objects, though 2), and to a lesser degree 3), involve subjects in an interesting sense (the classifications by reflectance spectra are relative to the way the human visual system works, and would have little interest in isolation from the latter). More importantly, it also applies to the order according to how immediate our acquaintance with the phenomenon is, i.e. how much epistemic trouble is required to get acquainted with it. There is a third sense in which the heading is misleading, however: the connotation of ‘being a more or less subjective property’ had better be avoided.

The ‘closer to or further from a human perspective’ ordering also gets things right for some levels in the *mereological* sense. As science progresses and gets increasingly technically advanced, we become acquainted with objects on a smaller and smaller scale. It takes more epistemic effort to get clear about quarks than molecules; in that sense, the latter are closer to a human perspective than the former.¹² However, this doesn’t hold across the board. In the mereological hierarchy, the human perspective is placed somewhere in the middle; as we discover planets, then solar systems, then galaxies, we move up the mereological hierarchy, but further away from a human perspective. So the ordering according to human perspective doesn’t get things completely right.¹³

Nevertheless, if what we seek is an explanation of the fact that what is really several orderings have been taken to be one, rather than a genuine sense in which the orderings are actually unified, the idea of ordering according to closeness to a human perspective seems to do the job. Whether we stick with the general formula or choose the one in terms of the amount of epistemic effort needed for acquaintance with the phenomenon, a partial explanation of the conflation of the various senses of levels is forthcoming: an ordering on this basis will yield a picture where the senses are mixed. In rough outline, closest to a subjective standpoint we find phenomenal and later functional levels, where the relevant relations will be supervenience, realisation and constitution. When we move further away, we find levels of physical qualities.

¹² The concept of atoms was in use before that of molecules. But this poses no problem for the suggestion; the ordering concerns the *referents* and not the concepts. Thanks to Sarah Broadie for raising this question.

¹³ Thanks to Katherine Hawley for pointing this out.

At the lower end of the scale, we get acquainted with smaller and smaller objects as science progresses, and here mereological levels are of primary importance. The higher end of the mereological scale does not fit the bill, but as most philosophers have been interested primarily in the lower end of the scale, with the expectation that macro-phenomena can be explained in terms of micro-phenomena, this might have been overlooked. With those phenomena that do fit in, we get something like a single ordering encompassing levels in different senses. This may be what has made people fail to distinguish the different senses of levels.¹⁴ (It may also explain Schaffer's puzzling claim that the different senses of levels have to comport in order to be in good standing.)

To sum up, there are different senses of levels that should not be conflated. This does not undermine the idea of levels, though it creates a potential for misunderstandings and connotations. There is a lot further work to do in distinguishing the different senses of levels and investigating them in detail; we have only scratched the surface. But this brief account should suffice for our purposes.

2. How to individuate levels?

When the idea of levels first became common currency, it was in the context of the positivist 'unity of science' hypothesis, on which physics was understood to constitute the basic level,¹⁵ and the higher levels of chemistry, biology, psychology, and social science were thought to be reducible to lower levels in a strict hierarchy – social science to psychology, psychology to biology, biology to chemistry, and chemistry to physics.¹⁶ Against this background the question of level individuation can be addressed by reducibility considerations. Putnam and Oppenheim distinguish six levels: elementary particles, atoms, molecules, cells, (multi-cellular) living things, and social groups. On the issue of levels individuation, they say:

¹⁴ E.g. Kim, in the quote earlier in the chapter; Kim has later changed his view on this, see e.g. 1998 p. 80-86, where he proposes to keep different senses of levels apart. See also Putnam & Oppenheim (1958).

¹⁵ Or levels, as on Putnam & Oppenheim's account.

¹⁶ See e.g. Putnam & Oppenheim (1958).

We maintain that each of our levels is necessary in the sense that it would be utopian to suppose that one might reduce all of the major theories or a whole branch concerned with any one of our six levels to a theory concerned with a lower level, skipping entirely the immediately lower level; and we maintain that our levels are sufficient in the sense that it would not be utopian to suppose that a major theory on any one of our levels might be directly reduced to the next lower level. (Although this is not to deny that it may be convenient, in special cases, to introduce intervening steps.) (Putnam & Oppenheim 1958)

In short, the way to tell what levels there are and to tell them apart is to consider whether each is 1) needed for reduction from higher to lower levels, and whether 2) it is likely that its theories can be reduced to those of the level below it.

The way I understand the notion of levels is different in at least two respects: 1) it doesn't come with a presupposition of reducibility, and – more relevant for present purposes – 2) it is better understood as something more fine-grained than the 'science-wide' levels of positivist thought. For example, while surface structures and reflectance spectra are best thought of as different levels and different location options for colours, they both belong within the domain of physics (and if it takes psychology and/or neuroscience to pick out the relevant classes of physical properties, this is the same for both).

What individuates levels in the more fine-grained sense we are after?

With questions like this, it might make sense to be deliberately vague about the some of the detail because the framework is most useful if people are allowed to fill it in according to philosophical preferences. How people will want to individuate levels will depend heavily on their views on the phenomenon under discussion (for example, some die-hard physicalists may want to deny the existence of the higher levels altogether), so fixing on a particular picture might rule out some of the positions in discussions for which the levels metaphor is helpful. So the best way to make clear the intended notion of levels may simply be to give examples of levels in different areas, leaving room for different ways to conceive of them. Also, a general recipe might be hard to come by because the differences between neighbouring levels seem to vary with the subject matter and the particular levels in question.

However, some general criteria of levels difference are discernible. First, different levels will usually come with *different criteria for detection* of the

phenomenon (phenomenal heat versus temperature measurements with mercury thermometers versus the methods to detect motion in molecules etc.). Secondly, level differences might be expected *where the phenomenon starts behaving like a different kind of property* – e.g. where something that seemed to be of dispositional nature starts looking like a categorical property,¹⁷ or when we cross the boundaries between phenomenal, functional and physical properties. Finally, *modal intuitions* will be important in distinguishing between levels; if two aspects of a phenomenon, though usually or always co-instantiated in the actual world, can come apart in counterfactual cases, there is a case for treating them as different levels (e.g. surface structures and dispositions to elicit colour experiences, or being a statue and being a lump of clay).

The issue of level individuation deserves closer scrutiny. But an intuitive sense of levels and level individuation, together with the rough-and-ready criteria listed above, should suffice for our purposes.

3. Ontological differences or just modes of presentation?

A crucial question about the levels framework is the question about its ontological implications. Is reality itself stratified? Are there really different levels of properties (, objects, states of affairs)? Or does the levels metaphor just cover different modes of presentation, different ways to view and talk about what is really one and the same phenomenon?

A first thing to note about this issue is that admission of ontologically different levels does not imply *reification* of levels, i.e. thinking of levels as something like shelves for the properties etc. to sit on – shelves that exist independently of the properties sitting on them. There is no need to presuppose the existence of extra entities besides the ones the location disputes are concerned with. Exactly as we can talk about e.g. points and lines in space compatibly with thinking that space is not an absolute and independently existing framework, but just a matter of relative distances

¹⁷ Or a relatively categorical property if you think that dispositionality/categoricity are a relative, not an absolute matter.

between things etc., we can talk about levels without thinking of them as reified. Level realism can thus be a fairly moderate realism.¹⁸

Secondly, we should note that the answer to our question should be *relativised to the different senses of levels* we have distinguished. It seems fairly uncontroversial that there are levels in the mereological sense – i.e. that there are things that stand in the relation of parts to wholes. There are ways to disagree with this claim, but none of them seem very plausible.¹⁹ However, the levels most relevant for our purposes are non-mereological ones, related by supervenience, realisation or constitution relations. Are there ontologically different levels in this sense?

It might seem that the levels framework is in trouble no matter what stance is taken on this issue. If level talk is invested with ontological significance, this could be thought of as question-begging against certain positions in the disputes that the framework is used to set up – most notably, identity theories. Also, it involves taking on metaphysical commitments that are unnecessary in the eyes of some (e.g. identity theorists), and would need justification. If, on the other hand, the level talk is not to be taken literally, but just concerns different modes of presentation, does it say anything at all?

The answer to this double challenge, I think, is the following: The levels framework provides a useful way of setting up discussions and making clear positions in many areas (including colours, philosophy of mind, and our other examples), in advance of settling the issue of what theory is appropriate for the area – and, in particular, without begging questions against identity theories. We might talk about e.g. states of brains and states of minds as ‘different levels’, and then ask about the relation between them, in a way that allows ‘Identity!’ as a perfectly legitimate answer. If that answer is chosen, we can conclude that the levels talk was just a way of speaking, and that the differences were just differences in the way the

¹⁸ Thanks to Katherine Hawley for helpful discussion of this.

¹⁹ There are at least the following ways to reject mereological levels: 1) the view that only atoms (i.e. smallest parts) are ultimately real, while aggregates are not, 2) the view that there is no privileged way of dividing things into parts, and that without natural groupings into e.g. molecules, atoms etc., the idea of mereological levels makes little sense, and 3) the view that any division into parts is artificial and should be avoided.

phenomenon presented itself to us. If, on the other hand, we choose a different answer (supervenience, realisation etc.), we might invest the levels talk with ontological significance, for a given case or across the board. No choice needs to be made in advance of settling a particular location dispute, and no uniform choice needs to be made for all cases.

Thus, both of the above challenges get something right, but also get something wrong. The answer to the second challenge is that even if levels talk can be used in a neutral way, it does not imply a commitment to remaining neutral. If the objector feels too little has been said, then she should await our account of the ontological level differences for particular domains. The answer to the first challenge is that levels talk can make sense without ontological level differences. The levels framework provides a useful way of setting up discussions prior to settling the location question, and is compatible with answers that rule out ontological level differences. When the stage is set by ontologically innocent levels talk, we can move on to the location issue itself, including the issue of whether or not there are ontological level differences at play.

Ontological implications of level talk, at least in the supervenience sense, need not be thought of as an all-or-nothing matter. The different supervenience relations make for different degrees of connection between the related properties, and correspondingly different degrees to which it is reasonable to talk about ontological independence. If the properties presented as occupying different levels are identical (in supervenience terms, related by symmetrical logical supervenience), then it is true in the strongest sense that in this case there are no ontological level differences. If they are connected by 'strong', logical supervenience, the ontological difference between levels is of a fairly lean sort; the properties are very closely connected, and the high level ones presumably determined by the low level ones. There is thus a sense in which the alleged high level properties are nothing over and above the supervenience base of low level properties. If the properties are related by weaker supervenience relations (mereological supervenience, or maybe weak or Blackburn-style supervenience), then the properties are different, though not entirely independent, and there are definitely ontological level differences. Their exact nature will depend on the supervenience relation in question. If the properties on 'different

levels' are entirely unconnected, talk of different levels stops making sense, as it becomes hard to see how entirely unconnected properties could be arranged in a hierarchy in a way that would make it sensible to conceive of them as different levels. Thus, while weaker connections make for stronger ontological independence between levels up to this point, no connection at all does not support the strongest version of 'levels with ontological implications'.

On the picture I suggest, the question of ontological implications should be settled together with the location dispute in each domain. A negative answer could be motivated by an identity theory (pain is C-fibre firing, heat is molecular mechanical energy, etc.). It could also be motivated by a primitivist view: the view that the disputed properties are *sui generis* qualities, and not reducible to, or locatable on the level of, properties of any other kind. As hinted at in Ch. 2, this position sits uneasily with a levels framework because the simple properties it postulates would seem to compete with the traditional location candidates; for example, the position would face causal exclusion arguments based on competition for causal work between the primitive properties and traditional low level candidates. So primitivists would have a motivation for rejecting the level framework as a false picture.

Some people deny the existence of different levels on quite general grounds, and not because of location preferences within a particular domain. For completeness, I shall mention a few such positions. One such view is that only one level of properties is real, and that all other alleged levels of properties are either 1) identical with these properties, or 2) reducible to them (in some sense strong enough to take away any ontological commitment to higher levels), or 3) are simply confusions. Usually, the privileged level is thought to be a basic level of physical properties. (Notice, however, that the basic level in the supervenience sense of levels could be an intermediate mereological level, if all smaller parts supervene on this intermediate level.)

For level optimists, a natural strategy against such views would be to examine their motivations and search for questionable presuppositions. If, for example, the view is motivated by causal exclusion arguments combined with a presupposition of a basic level, the level optimist could argue that the presupposition of a basic level is

unjustified, or that something is wrong with causal exclusion arguments. More on this in section 4.)

Another general level-sceptic's strategy is to argue that everything belongs on one level, even though none of the formerly assumed layers enjoy a special status. L.A. Paul argues for such a position (for objects; it is not clear what the story would be for properties). Her view is that objects are fusions of properties, and that presumed different level objects such as persons and their bodies stand in a relation of partial overlap. This means that there is no basis for a level ordering between them.²⁰

To sum up the results of this section, the answer to the question about the ontological implications of the levels framework is a rather complex one: Levels talk does not require reification of levels, and it can be used in an ontologically neutral way for setting up discussions. The next question is to what extent there are really ontological levels differences. The answer to this question must be relativised to the different senses of levels (mereological, supervenience etc.), and probably to particular domains (though there are views that are level-sceptic across the board). If the disputed properties are related by supervenience relations, we get a range of different degrees of dependence, and the weaker the dependence, the stronger the ontological level differences – but only up to the point of no dependence at all, where it becomes senseless to talk about levels. The existence of different levels can also be doubted on quite general grounds, e.g. causal exclusion considerations. In the next section, we shall examine a presupposition of many such views: the idea of a basic level.

²⁰ L.A. Paul: 'Against a Layered World', unpublished manuscript.

Paul's view is intended to solve the problem with causal exclusion. The idea is that the 'higher level' and 'lower level' object can both be doing the same causal work without over-determination because the causal work is done by the part (i.e. set of properties) that is shared between persons and their bodies. As causal exclusion problems arise for properties as well as objects, it would make sense for Paul to have a similar story to tell about those. But her paper is silent on this issue, and it is not clear how her story can be extended to cover properties as well as objects.

4. Is there a basic level?

In his paper 'Is there a Fundamental Level?' (2003), Jonathan Schaffer makes a comprehensive case that belief in a fundamental level is unjustified for all the different senses of levels. He introduces the idea of a fundamental level as follows:

'Talk about 'the fundamental level of reality' pervades contemporary metaphysics. The fundamentalist starts with (a) a hierarchical picture of nature as stratified into levels, adds (b) an assumption that there is a bottom level which is fundamental, and winds up, often enough, with (c) an ontological attitude according to which the entities of the fundamental level are primarily real, while any remaining contingent entities are at best derivative, if real at all. (2003, p. 498).

His paper takes issue with (b), the assumption of a basic level, and argues that it is unwarranted. The question of the existence of a basic level – and our motivation and justification for believing in one – arises for each of the senses of levels (mereological, supervenience etc.) that we've considered. It may be that the levels comport, but if they don't, there may be a fundamental level in some senses and not in others. For example, matter might be infinitely divisible compatibly with there being a fundamental supervenience base – say if the parts below a certain level are boring in the sense that they supervene on the wholes. So really there is not one but several questions about a fundamental level.

For levels in the mereological sense, the idea of a basic level is widespread, but as Schaffer points out, many scientists and philosophers have also thought about matter as infinitely divisible into smaller and smaller parts, and there is nothing a priori wrong with this picture.²¹ The question of the existence of a basic level is an empirical one.

For levels in the supervenience sense, the idea of an ultimate supervenience base for everything is presumably equally widespread, and might be correct even if mereological fundamentalism is wrong. But its correctness is not beyond doubt. For example, it would be wrong if supervenience and mereological levels comport, and higher levels supervene on ever-deeper levels on the infinite divisibility picture.

²¹ See Schaffer (2003), p. 501-502, where he considers and rejects three a priori arguments in favour of mereological fundamentalism.

How about ‘levels’ related as roles and realisers? The traditional functionalist picture described in section 1 needs the presupposition of a basic level. Higher ‘levels’ or ‘orders’ of properties are properties of having certain properties of another kind: real properties that can do causal work. And while some role properties may be realised by further role properties, there must be a basic level where the instances that make true the implicit quantifications in the role properties are found. It might perhaps be possible to construct a version of functionalism (or something close) that will work without the presupposition of a basic level, but it would have to be very different from functionalism as traditionally understood.

For levels of objects related by the constitution relation, the issue of a basic level does not seem crucial; here, the significant distinction is between one level and more than one level (i.e. more than one object to a place).

In the following, we shall concentrate on the two first senses of levels: the mereological and supervenience senses.

Schaffer (2003) argues that the presupposition about a basic level is unwarranted both in the mereological and the supervenience versions. The main problem with level fundamentalism in both versions is that unwarranted empirical presuppositions are made. In the mereological case, Schaffer takes the reasoning to be something like this (these presuppositions are the only ones that would warrant the fundamentalist’s claims):

- 1) There will be a complete microphysics,
- 2) the complete microphysics will postulate particles, and
- 3) these particles are the mereological atoms.²²

In the supervenience case, similar presuppositions are needed:

- 4) There will be a complete microphysics,
- 5) it will describe a comprehensive supervenience base, and
- 6) this will license supervenience-only fundamentality.²³

²² Schaffer (2003), p. 502.

²³ Schaffer (2003), p. 510.

In both cases, the reasoning is flawed, simply because we don't have a reason to believe that there will be a complete microphysics. Science might continue to make progress without ever reaching a point at which the job is finished (or it might end or go astray before the goal is reached). 2) and 3) are also unjustified, as there is no telling what a complete microphysics will say until we've discovered it. Scientific development is notoriously unpredictable, and there is no guarantee that we have got the basic terms (e.g. particles) right yet. Also, many times scientists have believed we have found the 'atoms', only to discover that these, too, have parts; if history suggests anything, it is that the idea of smallest particles is a presupposition of ours; it is certainly not something we have empirical reasons to believe in.²⁴

For the supervenience case, the situation is a little better, as 5) and 6) are more plausible than 2) and 3). According to Schaffer, what defines a complete microphysics is that it tells a complete causal story, and a complete causal story will also serve as a complete supervenience base.²⁵ So 5) is warranted given 4), and is correctly neutral on what a complete microphysics would say. And 6) follows from 5); if we have a comprehensive supervenience base, then this licenses fundamentalism in the supervenience sense (but not in the mereological sense; the ultimate supervenience base might well be an intermediate mereological level). Supervenience-only-fundamentalism thus depends on fewer doubtful assumptions than mereological fundamentalism: all that is needed is the presupposition about a complete microphysics, while the view can remain (correctly) neutral on the basic terms of that completed science, and on the issue of infinite divisibility. But still, both versions depend on an unwarranted assumption.²⁶

²⁴ This paragraph is based on Schaffer (2003), p. 503-5.

²⁵ One might object that causation, like particles, might be one of the components that do not make it into completed science; who knows what completed science will be like? Dialectically, this does not damage Schaffer's argument, since it would make it even harder for the proponent of supervenience-only fundamentality to prove her point. Also, from the perspective of our current knowledge, the 'full causal story' component seems more likely to make it into completed science than the idea of fundamental particles. So supervenience-only-fundamentalists are still better off than mereological fundamentalists.

²⁶ Schaffer (2003), p. 510-12.

Schaffer concludes that we should distinguish three different metaphysical pictures: infinite divisibility (the negation of mereological fundamentality), supervenience-only-fundamentality, and full-blown (mereological plus supervenience) fundamentality. If we keep discovering novel structures on deeper levels, that would constitute defeasible evidence for infinite divisibility. If we discover a complete microphysics, supervenience-only-fundamentality will be justified (in Schaffer's view, this is the only 'expensive' empirical assumption supervenience-only-fundamentality requires). And if that microphysics postulates mereological atoms, then full-blown fundamentality will be justified. But until then, we have no reason to assume that there is a basic level.

Schaffer comes close to suggesting that, given no guarantee that there will be a complete microphysics, we should suppose infinite divisibility ('if we keep finding new levels with novel ontological structures, then that suggests infinite divisibility'). But given our current state of knowledge, agnosticism seems the proper attitude to the question of a fundamental level.

The issue deserves a lot more attention. With the exceptions mentioned, I find Schaffer's arguments convincing. But this brief discussion is aimed mainly to show that the presupposition of a basic level, in the supervenience or mereological versions, requires further justification.

4.1 Level fundamentalism and high-level locations

The question of a fundamental level is dialectically important when assessing the potential of theses of high-level location, including response-dependence theses. As Schaffer points out, level fundamentalism naturally gives rise to an attitude of inequality between levels. The idea of a basic level with a privileged ontological status might cast doubt on the importance of higher levels altogether. In the light of this picture, higher levels might be thought of as non-existent by one of the routes considered in the previous section (identification, reduction, or elimination). Or they may just be thought of as somehow less real and less important than the basic level. This would make theses of high level location, e.g. response-dependence theses, a lot less interesting than they might have seemed.

More concretely, the idea of a fundamental level plays a crucial role as backdrop for two standard argument types for low level locations: Arguments from ontological economy and causal exclusion arguments.

Reduction of one level of properties to another might seem to have a pay-off in terms of ontological economy even without the presupposition of a basic level. Reducing one class of facts to another is an achievement no matter what happens with other types of facts. However, if there are infinitely many levels, the achievement will be a lot less significant than if we are dealing with a limited number of levels; one step down an infinite road doesn't do much good. If, on the other hand, we can reduce the levels one by one with the prospect of reducing everything else to a single level of well-understood facts, there is a real point to the reductions. This is how arguments from ontological economy depend on level fundamentalism.

Causal exclusion arguments rely even more heavily on the idea of a basic level. They entail that any level of properties or objects will be causally superfluous if there is a lower level to which the causal work can be attributed. The line of thought can be summed up as follows:

1. The disputed phenomenon (say, mental states) can cause low level events,
2. any low level event has a sufficient low level cause, and
3. no causal over-determination is at play,
4. so a low-level location should be chosen over a high-level location that would render the disputed phenomenon causally impotent.

If this argument works, then high-level locations will be incompatible with causal efficacy, and so will often be unattractive. However, if there is no basic level, the outcome may be very different. If the hierarchy of levels continues down ad infinitum, and if the argument can be repeated for each level, transferring the causal powers to the level below it (as Block and others have argued²⁷), it would seem that 'the causal powers would drain away down a bottomless pit' (Kim's phrase).²⁸ No level would be doing the causal work. So either there is something wrong with the

²⁷ E.g. Block (1995).

²⁸ Kim (1998) p. 81. See Kim (2003) for a defence against the causal drainage problem.

argument, or there has to be some principled reason to stop the regress at some point, or causation must be an illusion.

An advocate of the causal exclusion argument might take the problem with causal drainage to show a priori that there must be a basic level. After all, we are pretty certain that there is causation going on in the world, and if the idea of infinite descent would conflict with this commonplace, then so much the worse for the idea of infinite descent. This just shows that there must be a basic level to hold the causal powers. This line of thought (along with other a priori arguments for the existence of a basic level) strikes me as misguided. The question about the existence of a fundamental level (in every sense of levels) is presumably an empirical one, as Schaffer points out, and this kind of a priori reasoning seems the wrong kind of evidence for a conclusion on the matter. If the causal exclusion argument depends on the presupposition of a basic level as an a priori matter, this suggests that something is wrong with the argument, even if we can't say exactly what it is.

The issue of causal exclusion and drainage deserves a longer discussion that, again, we shall have to cut short.²⁹ The important point for our purposes is that the presupposition of a fundamental level plays a crucial role in arguments for low level locations. If this presupposition is flawed, that makes high level locations such as response-dependence theses a lot more attractive. Schaffer has made a convincing case that the presupposition of a basic level is a lot less secure than people seem to think. So to the extent that levels fundamentalism fuels scepticism about high level locations, that scepticism seems unwarranted.

5. The levels framework and the nature of properties

What presuppositions, if any, does the levels framework make with respect to the nature of properties? And what implications, if any, does the levels-based approach have for the way we should think about properties?

Recall our distinctions from Ch. 1 between criterially and referentially governed concepts, and between substantial and deflationary properties. A referentially governed concept is intended to refer to an independently existing,

²⁹ See Block and Kim for details; e.g. Block (1995) and (2003), Kim (1998) and (2003).

substantial property, and to have its extension determined by the extension of that property. By contrast, a criterially governed concept applies to everything that satisfies certain criteria, whatever its nature.

On a deflationary conception of properties, instances of a property need have nothing in common apart from being classified together, whereas the instances of substantial properties must have something in common in virtue of their own nature. We might think of all properties in one or other of these ways, or combine them into a mixed view on which some properties are substantial and some merely deflationary.

The levels framework is naturally combined with a mixed view of properties. Moreover, it suggests a picture on which *substantiality is a matter of degree*. Such a picture is attractive independently of the levels framework. If the idea of substantial, natural properties is understood as properties visible from ‘the view from nowhere’ and relative to no cognitive and perceptual apparatus, then there may not be many of them, and most of the concepts we presently have presumably do not track such properties. But it would seem too quick to conclude that most of our concepts are only correlated with merely deflationary properties. There seems to be a world of difference between concepts that are determined only by consensus or by deferring to a group of trend-setters (e.g. the concept ‘U’ mentioned by Pettit,³⁰ which applies to whatever the members of the Sloane Square set agree that it applies to), and a low-level scientific concept that is close to getting things right but still holds a little relativity to a human perspective (and from each of these to the intermediate case of everyday concepts like ‘table’ or ‘hill’). Classifications may be due partly to us and partly to features of the environment, and so the extension of a concept might be determined in a way that does not completely fit any of the two paradigms. A natural way to explain the situation is to recognise intermediate steps, or degrees of substantiality.

This picture makes a lot of sense if we think of our enterprise of getting to know the world in the following way: we start by making classifications that are highly relative to a human perspective, then try to move further away from our own

³⁰ Pettit 1991, p. 611.

perspective and ‘correct’ for its influence, and to discover categories that would make sense even from a perspective very different from our own. But there is no guarantee that there is a final set of categories that are completely perspective-neutral, and which we might reach if only we looked deep enough. Correspondingly, it seems wrong to require that our concepts must either refer to those final properties or not refer at all.

The levels framework – in the intuitive version described under ‘a unified sense of levels’ above – offers a natural way to understand such degrees of substantiality, and how they come about. The most substantial properties are the ones found on lower levels, far away from a human perspective, and requiring a lot of epistemic trouble for acquaintance. The less substantial properties are the ones found on the higher levels. The framework thus comes with a way of thinking about properties that suits not only its own requirements, but also intuitions that are independent of it.

The levels framework makes most sense if we think of substantiality as a matter of degree. If only (completely) substantial properties are recognised, location debates become impoverished. Each location dispute will have very few levels of relevant properties, and often none at all; for colours, for example, none of the candidate properties will qualify as properties in the substantial sense. The situation is a little better if we take a mixed view and allow for deflationary properties as well, as this would give us more candidate properties to choose from. But on a deflationary conception of properties, location disputes are not very interesting; what would be under discussion will presumably be which set of criteria governs the concept under discussion, where the candidate deflationary ‘properties’ are each associated with a set of candidate criteria. This sort of dispute would not be totally uninteresting, but would be less momentous than a dispute over the location of the reference of a concept. If substantiality is considered a matter of degree, we get location disputes with the number of location options that we would have expected, and with something more at issue than the criteria governing the concept.

A mixed view, combined with substantiality in degrees, is thus a natural choice given the levels framework. However, even if you don’t accept such a view, the picture can presumably make sense. If you believe in deflationary properties only,

you could talk about properties all the way up, and re-describe the low level properties as something like ‘properties helpfully describable in terms of scientific concepts’ or similar. If, on the other hand, you believe in substantial properties only, you may say that there are properties on the lowest levels, maybe complexes describable in terms of (combinations of) properties on intermediate ones, and presumably nothing ontologically interesting on the highest levels; ‘high level’ concepts fail of reference, and are either criterially governed or hopelessly flawed.

In the following chapters, I shall mostly talk as if there are properties of both kinds – substantial properties and (merely) deflationary properties. If you don’t like this way of speaking, please feel free to translate my claims into your preferred terminology using the paraphrases suggested above.

6. Justification and applications

Why adopt a levels framework? The best way to justify the framework is probably to show some of the interesting uses it can be put to: setting up discussions, suggesting solutions, clarifying positions and their similarities and differences, and exposing equivocations and fallacies. The prospects for interesting results seem good in many areas and many respects. We’ve already seen how the levels framework can be useful in setting up the discussion about the nature of colours. It also provides a clear perspective on the nature of mental states, and of the many components in our conception of these. The advantages of keeping these clearly apart are evident in Chalmers (1996). Dispositions are another area where the levels-based way of thinking might prove useful. I suspect that a lot of confusion can be cleared away by clearly distinguishing dispositions-as-response-patterns and dispositions-as-bases. I hope to develop this further in another context, and will only briefly touch on these issues in this work (see Ch. 6). If the idea of levels is applied to objects as well as properties, it also provides interesting results with puzzles of material constitution. In the appendix below, I shall explore a levels-based solution to a range of such puzzles, e.g. that of Theseus’s ship.

Last, but not least, the levels framework can help us get a clearer perspective on the debate on response-dependence. By the end of this work, it will hopefully have proved its worth by helping to shed light on this hard issue.

6.1 The levels framework at work for response-dependence

What can the levels framework do for response-dependence?

First and foremost, the framework provides a clear perspective on the disputes in which response-dependence theses are one of the live options. It offers a way to gain an overview of the various intuitions and other ‘data’ in the debates and the weighting between them. It thus helps in setting up the problems and stating the demands that acceptable solutions must meet, making clear what ‘data’ must be accommodated or explained away. This advantage of the levels framework is more relevant with particular examples of location disputes. So unfortunately it won’t be brought out very clearly in what follows, as we shall focus on general issues regarding response-dependence. But it will be important in subsequent application of the results to particular domains.

A second advantage of the levels framework is that it can help to pin down what exactly the various sorts of response-dependence theses say, and help distinguish between the various versions in the literature and in logical space. In particular, it is useful for stating the difference between response-dependence theses of subject matter (such as Johnston’s and Wright’s versions), which entail a high level location of the disputed phenomenon, and response-dependence that concerns concepts only (Pettit’s version), which is neutral on the location issue.

Thirdly, the levels framework helps to bring out the potential of the various versions of response-dependence. The second half of Ch. 5 is devoted to the advantages of combining Pettit-style response-dependence with a levels-based approach. In Ch. 6, I shall argue that the levels-based approach offers new ways to tell response-dependent from -independent subject matters – ways that are less problematic and easier to handle than the traditional criteria. The levels framework also provides a clear perspective on some further issues related to response-dependence, e.g. the missing explanation argument (which I shall say a little about in Ch. 4, though a more thorough discussion must await a later occasion).

A fourth advantage is that the levels framework provides a way to expose prejudices against what in this framework will be high level locations – for example, the ill-founded belief in a basic level with a privileged status. It thereby paves the

way for theses of high-level location, including response-dependence theses of subject matter.

Finally, adopting the levels framework makes it easier to argue for theses of high level location such as response-dependence theses because it allows for different kinds of properties with different degrees of substantiality and objectivity. On a picture according to which there is a limited class of privileged, real properties – perhaps those of completed science, and in any case properties that in the level terminology belong on low levels – the prospects for response-dependence would be dim, and the consequences of a response-dependence thesis very anti-realist. But if there are several levels, and especially if there is no guarantee that a basic level exists, response-dependence theses will be more plausible and less radical.

7. Summary

Though the issues discussed in this chapter call for further investigation, I hope the discussion has conveyed a sense of the levels framework and its potential and problems. Let us briefly sum up our results.

The relationship between properties, objects or states of affairs on neighbouring levels can be supervenience, realisation, constitution, and – less relevantly for our purposes – mereological relations. If the levels talk is not invested with ontological implications, it can also be identity. Different relations may be relevant in different contexts. Levels related in these different ways may have been conflated because they largely fit into a common ordering according to closeness to a human perspective (though with exceptions). In the following, we shall focus mainly on levels related by supervenience relations, though an investigation into the exact nature of these will not be attempted.

The prospects for a general recipe of level individuation seem meagre. But we can distinguish some rough-and-ready criteria, including differences in detection criteria, modal intuitions, differences in kinds of properties, etc.

The levels apparatus can be useful with or without ontological implications. If kept ontologically neutral, it is helpful for setting up discussions of location issues, and depending on the stance taken in these disputes, it can be invested with ontological significance.

The presupposition that there is a basic level is widespread, but ill founded, at least in the supervenience and mereological versions. The proper attitude on the matter is agnosticism. This makes high level locations more attractive than they might have seemed, as the argument types most commonly used against high level locations – arguments from ontological economy and causal exclusion arguments – depend heavily on the presupposition of a basic level.

The levels framework makes it natural to think of substantiality of properties as a matter of degree. Such a view is attractive on independent grounds, and the levels framework offers a natural way to understand and justify it. The levels framework itself is best justified by its usefulness in setting up discussions, suggesting solutions, clarifying positions and the relationship between them, and exposing equivocations and fallacies in various areas. Hopefully, the following chapters will demonstrate its potential with respect to response-dependence. On the next pages, we shall see how it might be useful for solving puzzles of material constitution.

8. Appendix: Theseus's ships³¹

In this appendix, I apply the levels framework to the puzzles of material constitution, and propose a levels-based solution to some of them. The purpose is to show how the idea of levels can do useful work outside the area of response-dependence. First, I'll introduce the puzzles, basing the exposition on Michael Rea's (1997) introductory article. Then I'll sketch the levels-based solution, and finally discuss some loose ends, objections, and advantages of the proposal.

8.1 The puzzles of material constitution

Puzzles about material constitution take many forms. One of the classics is the puzzle about Theseus's ship. Theseus's ship sails the seas, exposed to wear and tear, and all its parts are gradually replaced, until not a single original part is left (for convenience, let us adopt Rea's anachronistic version and say that the new parts are made of aluminium). The original parts are collected by a historian, reassembled into a ship, and placed in a museum. Now, which ship is identical to the original ship: the aluminium ship sailing the seas with the same crew on board, or the museum ship that was built of the original materials? Both have some claim to be identical with Theseus's original ship. But they are certainly different from each other. So by transitivity of identity, they can't both be identical to the original ship, unless we want to say, contrary to intuitions, that there were two ships from the beginning.

This is just one puzzle of material constitution; there are lots of others. One is Alan Gibbard's puzzle about the relation between a statue and the lump of clay it is made of, both of which exist for the same amount of time in exactly the same place, whereas they differ in their modal properties. For example, the statue could survive the replacement of one of its hands, whereas the lump couldn't, and the lump, but not the statue, could survive being re-shaped into something else. A third puzzle concerns the cat Tibbles, who loses her tail in an accident. After the accident, the cat-

³¹ Thanks to Katherine Hawley and the audience at a research seminar in St. Andrews, particularly Patrick Greenough, Crispin Wright and Daniel Nolan, for helpful discussion of this material.

part Tib – Tibbles minus her tail – and Tibbles occupy the same space. Should we say that Tib is now a cat? If so, wasn't she one before? In that case, how many cats were there before? (For example, what about the cat-part consisting of Tib minus two hairs? Etc.) There are lots of other similar puzzles, but these should suffice to give an idea of the issues under discussion.

8.2 Rea's recipe for puzzles of material constitution

Despite their differences, the puzzles are built from similar materials. Rea (1997) extracts a useful general formula. He reconstructs all the puzzles as arising from the conjunction of five individually plausible premises of characteristic types.

The first is an *Existence Assumption* (EX in the following): Each puzzle assumes that there is at least one object of a certain kind, and that this object has parts. In the case of Theseus's ship, this object is a ship built of planks etc. In general, there will be an assumption of the form 'There is an F, and there are ps that compose it', where the F is a sortal (ship, statue, cat etc.), and 'ps' refer collectively to the parts that together compose the object.

The second assumption is an *Essentialist Assumption* (EA): Each puzzle assumes that there is some relation r such that whenever we have an object of the kind postulated by the existence assumption, we have something that essentially bears this relation to its parts. An example could be this: if we have a ship like Theseus's, we have something that essentially contains at least a certain proportion of the planks it is made of. (Note, however, that the essential relation could be chosen very differently; for example: when we have a ship, we have something to which it is essential that its parts are organised in a whole that has a certain causal history, continuity with the original ship, or whatever.)

The third assumption, *the Principle of Alternative Compositional Possibilities* (PACP), is complementary to the Essentialist Assumption. It says that the parts of the F compose something that does *not* essentially bear the relation stated in (EA) to its parts. In our example, the assumption will be that the parts of our ship compose something that could have all its parts changed without ceasing to be. (Or, corresponding to our alternative formulation of the essentialist assumption: the parts of the ship compose something which could have had a different causal history etc.)

For each puzzle, there will often be many alternative choices of EA and corresponding PACP; any plausible choice, consistently applied, will be sufficient to generate the puzzle.

The fourth assumption is *the Identity Assumption (IA)*: For any objects x and y, if x and y share all of the same parts at the same time, then x is identical with y.

The fifth is *the Necessity Assumption (NA)*: For any objects x and y, if x is identical with y, then it is necessary that x is identical with y.

Combining these assumptions, we have 1) that there exists an F composed of ps, 2) that whenever we have an F, its parts compose something, call it a, which essentially bears relation r to its parts, and 3) that they also compose something, call it b, which does not essentially bear r to its parts. Since the ps compose both a and b at the same time, the identity assumption entails that a is identical with b. The necessity assumption takes us from this to the claim that it is *not* possible for a to be distinct from b. But since a essentially bears r to its parts whereas b can exist and fail to bear r to its parts, it follows that it *is* possible for a to be distinct from b. So we have a contradiction. All five assumptions, though prima facie plausible, can't be true together.

Rea distinguishes various ways of solving each puzzle by denying one of the five assumptions. Denying the identity or the necessity assumption will solve the problems once and for all, allowing no instance of the argument to go through. The appropriate versions of the three other assumptions can be rejected for particular puzzles.

Denying the identity assumption amounts to saying that there are really two distinct things, a and b – in the case of Theseus's ship, for instance, there are really two 'ships', and thus neither is *the* ship of Theseus. Such claims, which Rea terms coincident entities solutions, are usually backed by a story about entities of different kinds. The solution I shall explore is of this sort.

8.3 Level differences in puzzles of material constitution

The suggestion is that ships, like colours etc., can be viewed as a multi-level phenomenon, and that we should distinguish between a high-level, functional criterion for being the same ship, and a low-level, material one. The corresponding

‘high level ship’ and ‘low level ship’ can be seen as two different objects, though they can occupy the same space at the same time and count as one object for practical purposes.

In the case of Theseus’s ship, we have two conflicting intuitions about the original ship:

- 1) the ship can have its parts replaced without ceasing to exist
(this points to the aluminium ship as identical to the original)
- 2) the ship can’t survive a replacement of a certain proportion of its parts
(this points to the museum ship as identical to the original)

This pattern is familiar from location disputes where two or more sets of intuitions point to different locations of the disputed phenomenon, e.g. the two classes of intuitions about colours mentioned in Ch. 2. In the case of the ship, one intuition follows (what is naturally thought of as) a lower level phenomenon, the collection of parts (planks or elementary particles or whatever), and takes this to be essential to the object under discussion. This points to the museum ship as identical to the original ship. Another intuition follows (what is naturally thought of as) a higher level phenomenon, individuated by its causal history or something similar, and makes this the essential thing about the ship. Being a ship in this sense is probably a functional property. This points to the aluminium ship as identical to the original ship. When the two ‘ships’, the higher level one and the lower level one, part company in a way that gives equal support to the two intuitions pointing to different levels, we get a conflict between the two intuitions; this creates a ‘paradox’ of material constitution.

The concept *ship* (and the criteria for being the *same* ship) thus equivocates between both levels – i.e. contains material that points to two ‘locations’ – and could be construed as a mixture between two purified, or more systematic, or contradiction-free, concepts of ship: one that would clearly identify the museum ship with the original one (location with a low-level, material object), and one that would identify the aluminium ship with the original (location with a functional object).

(But it would probably be a mistake to try to purify the concept by discarding one of these components; the concept works fine for practical purposes – and presumably better than any of the single-level concepts would – and only gives

trouble in these few abstract cases. In this case, an equivocation view seems the natural choice.)

The outcome regarding the puzzle of Theseus's ship is that both ships are identical with the original given one of the disambiguations above: The museum ship is the same low-level, material ship as the original one, and the aluminium ship is the same high-level, functional ship as the original one.

If we take this route, we can say that there are really two ships in the beginning: a low-level, material ship, and a functional ship. These two can't be identical with each other, since one is identical with the museum ship, and the other is identical with the aluminium ship, and these two are clearly different. But saying this offends against the intuition that there can't be more than one material thing (completely) occupying the same space at any one time. We can label this intuition the spatial exclusion principle:

(SEP) For any two objects A and B, if A and B completely occupy the same spatial region at the same time, then $A = B$

I suggest that SEP, as stated, is too strong: It should be restricted to hold within a given level, but not across different levels. That is, it should be

(RSEP) For any two objects A and B *on the same level*, if A and B completely occupy the same spatial region at the same time, then $A = B$

This restriction to SEP is not just a theoretical construction, but seems to be recognised in ordinary practice. While we would say that there can't be two human bodies occupying the same space at the same time, few would see it as a problem that you and your body share the same space (and this is hardly because you and your body are identical with each other; your body could 'survive' your death as a corpse, while, well, you could not). The restriction is recognised where the levels in question are conceived of as different. Or better: the restriction is recognised *insofar as* the levels in question are conceived of as different. The more different we believe the levels to be, the more obvious it will seem that different level objects do not compete for spatio-temporal location in the way same-level objects do.

8.4 The suggested solution in Rea's terminology

In Rea's terminology, the sketched solution would amount to the suggestion that the Identity Assumption should be restricted in the same way as SEP above. The IA says that

(IA) For any objects x and y , if x and y share all of the same parts at the same time, then x is identical with y

The suggested solution is not to deny this, but to limit its scope: (IA) is correct only within a level. The revised identity assumption will be this:

(RIA) For any objects x and y *on the same level*, if x and y share all of the same parts at the same time, then x is identical with y

This suggestion needs backing by an account of levels. In particular, it needs to be backed by an account of level individuation. In the preliminary remarks about level individuation earlier, I mentioned diverging modal intuitions as a guide to level individuation. More needs to be said, of course. But suppose for a moment that the work is already done. Then we have the ingredients for the rest of our recipe for avoiding paradoxes of material constitution. For in that case, *whenever the EA and the PACP hold together, this shows exactly that two levels are in play*. For example, if EA says that the ship has some proportion of its parts essentially, this locates the (property of being the) ship at a low level – that of being an aggregate of particular parts. The corresponding PACP will say that it could survive the loss of these parts, and hereby locate the disputed phenomenon on a higher, functional-sounding level, and similarly for other cases and other choices of EA and PACP.³² So the five assumptions in Rea's taxonomy can't all be true together; whenever the EA and PACP are both true, that shows that two different levels are in play, and hence that IA does not hold for the case in question.

³² Rea's application of the 5-assumption taxonomy to some of the examples shows that the EA does not always have to follow the low level property, and PACP the high level one. If we formulate the essence in EA in a functional way, such as 'essentially consisting of parts (whatever they are) that are arranged cat-wise', EA will go with the high level property, and the corresponding PACP will go with the low-level aggregate of (accidentally) cat-constituting molecules.

8.5 Transitivity-of-identity puzzles without level differences

The suggested solution has an obvious limitation: It will only work if the two candidate objects can be reasonably thought of as belonging to different levels. But we can get problems for same-level objects that are very similar to puzzles of material constitution. For example, suppose we take two cars from the same production line, Carl and Carla, and start swapping their parts until it is no longer clear which of the cars is identical with the original Carl.³³ Both will have some claim to be identical with Carl, but both can't be, as they are clearly different from each other. In such cases, the levels-based suggestion will not help, as no level difference is in play. So for same-level cases of problems with transitivity of identity, we need a different account, presumably a story about indeterminacy.

Is this a problem for the levels-based suggestion? I think not. It would be a major problem if the supplementary story brought in to deal with same-level cases would work for different-level cases as well, since it would then make the levels-based solution superfluous. But it seems fairly clear that the problem with Theseus's ship (and Tibbles, and the statue) is *not* indeterminacy. This case is very different from the car case in that totally different considerations make the two ships seem to qualify as identical with the original ship, and depending on which set of criteria we consider as essential for shipness, one or the other will 'win'. For the cars, by contrast, their claim to be identical with Carl is based on the same sort of considerations: their containing such and such a proportion of its original parts (or a high level story could be chosen; the important point is the symmetry, not the level considered). For such cases, a story about indeterminacy seems to be the right response. The need for this supplement poses no problem for the levels-based solution; it just means that the levels-story will not solve all problems – not that it won't solve those it was designed to solve, or that it won't be necessary for dealing with these.

³³ The car example is due to Crispin Wright.

8.6 Is it too easy to claim level difference?

The cases of same-level, indeterminacy-based puzzles about transitivity of identity can actually help address a problem for the levels-based suggestion. The problem is that the account might seem circular. I have offered no independent account of how to recognise level differences when you see them; all I have said is that differences in modal intuitions, such as those driving EA and PACP, indicate level differences. If all it takes to generate a level difference is intuitions about different essences, and if level differences entitle us to plead failure of the identity assumption, then indeed there is no problem with puzzles of material constitution that fit Rea's recipe; the five premises will never be true together. But the solution might seem too easy if what entitles us to claim level difference, and hence failure of the identity assumption, is exactly that EA and PACP both hold. The account would stand a lot stronger if an independent account could be given of what individuates levels, and what it takes to generate a level difference. However, as we saw in section 2, a general recipe of levels individuation is difficult or maybe impossible to obtain. Our discussion centred on properties, but the task does not seem easier for objects.

Cases like the puzzle about Carl and Carla might help with this question. While it is hard to say what individuates levels, we have fairly firm intuitions about what *sameness of level* consists in. In the car case, it is clear that the puzzle does not hinge on a level difference, and that is why indeterminacy considerations are the natural place to look for a solution, whereas they seem irrelevant to cases like Theseus's ship. Where indeterminacy considerations seem relevant, then, it is likely that no level difference is at play.³⁴

The best thing to say might be this: In puzzles about same-level objects, indeterminacy considerations should be expected to do the job, and where they seem irrelevant, we have a reason to believe that a level difference is at play, and a reason to endorse the levels-based solution. In each case, you can pick the strategy that seems fitting for the case in question. However, it seems evident that often the latter will be the best choice.

³⁴ Thanks to Katherine Hawley for this suggestion.

8.7 Levels of objects or just levels of properties?

A third line of resistance to the levels-based solution to puzzles of material constitution would be based on the thought that the idea of different levels of objects is far too radical. One reason to think so would be the thought that while the nature of properties may be a bit of a mystery to us, at least we know what objects are. So while we can accept that different properties of different kinds (and maybe even different levels) are connected with e.g. redness, the idea that objects, too, are stratified in this unexpected way seems unpalatable. Another, more serious problem is the clash with the intuition that only one object can (completely) occupy a given spatial region at a given time (whereas properties do not exclude each other in this way).

In response to this worry, there are two overall lines to take. One would be to reformulate claims about different levels of objects as claims about different levels of properties had by the same object. For example, rather than thinking of a statue and the lump of clay it is made of as two different objects, we should say that the same object has the property of being a statue and also the property of being a lump of clay. Likewise, Theseus's ships will have the property of being a functional ship, and the property of being a material ship, but will still be one object.

This strategy would require a different solution to the puzzles of material constitution from the one suggested, though it could be based on similar materials. The suggestion could be that identity is relative, not to levels of objects, but to sortals, or perhaps to different sets of criteria for continued existence, corresponding to different level components in the concepts. So we could say that the aluminium ship is identical to the original ship qua functional ship, or according to functional criteria for being the same ship, while the museum ship is identical to the original qua material ship, or according to material criteria for being the same ship. In this way, all talk about levels of objects could be avoided. However, the suggestion raises problems of its own.³⁵

Another option would be to say that the ship has two levels of relevant properties – a high-level property of functional continuity, and a low-level property

³⁵ For sortal relativity solutions, see e.g. Geach (1980).

of consisting of particular parts – but that it can have only one of them essentially. The levels framework would help to separate these clearly and make it clear that we have to choose; we must decide which one is essential, and tidy up our concepts accordingly. This amounts to denying one of the relevant EA and PACP, and giving up the intuitions supporting it. (We could also choose a multi-level criterion for being the same ship, e.g. that the object must consist of at least 50% of the original parts *and* (*or*) having a certain amount of functional continuity. But this, too, raises problems. If the multi-level criterion is a *conjunction* of necessary conditions on different levels, we may – and, in this case, will – get the unwanted result that none of the ships qualify. If it's a *disjunction*, we get the original problem that both ships qualify.)

The other line would be to stick to the story about two or more levels of different objects. The main problem with this line is multiplication of the number of objects; there would be two ships where we thought there was one. A way to answer this worry would be that object individuation is a less accurate and more pragmatic matter than the objection assumes. For ordinary purposes, the functional and the material ship, or the statue and the clay, count as one object, probably because they occupy the same spatial region at the same time. But this does not prevent us from distinguishing more objects in a less everyday and more technical sense. In this sense, it can be correct to say that there are two objects. The situation is just as it is with properties: we have different candidates that coincide in some (ordinary) circumstances, and that are therefore conflated. This doesn't mean that there are several objects in the everyday sense, just as the levels equivocation found in colour concepts does not entail that my uniformly black jumper has several colour properties.

Both of these lines may be defensible. Though they sound like very different theories, they may both be legitimate ways of describing the same reality. However, investigating this matter further would take us too far afield.

8.8 Are ships material objects?

Suppose we adopt the latter strategy and conclude that Theseus has two different but spatially coinciding ships: a low level one that consists essentially of its parts, and a

high level one with a functional essence. Should we say that both are material objects, so that two material objects share the same space? Or should we conclude that only the low level ship is a material object, and that cats, ships and statues in the other sense are immaterial?

I think the answer is ‘yes and no’. In a stringent, technical sense in which a material object is nothing over and above its parts (and so could not survive changes in them), no; the functional ship obviously cannot satisfy this criterion. In a more relaxed, pragmatic, everyday sense, yes: ships, statues, etc. are material objects, even if they do have a functional (component to their) essence. Our everyday way of individuating material objects allows for changes in their parts; things grow, replace their cells, lose their tails or have their planks replaced. So we do not in practice require that something must have its parts essentially in order to be a material object. So there is nothing wrong with calling ships material objects, even if the story we have told is true.

8.9 The puzzles at work for the levels framework

Whether or not a levels-based solution to puzzles of material constitution works, the puzzles can be employed to support the levels framework. Higher level properties were supposed to be spooky and alien by comparison to the nice, familiar, well-behaved class of material objects, to which, therefore, everything else should be reduced if possible. But the puzzles show that material objects are not that well-behaved after all. They generate paradoxes – or at least they do so until we distinguish different levels within the class of objects normally taken to be material ones. So they can’t be used to set a standard for well-behavedness that higher level properties cannot meet.

If the problems with material constitution can be solved by introducing level distinctions within the class of material objects, this, of course, speaks strongly in favour of the levels framework. This – successful applications – may well be the best justification the framework will get.

8.10 Conclusion

Puzzles of material constitution display many of the same features as the location disputes (colours etc.) for which the levels framework was developed. Concepts like ‘ship’ can usefully be thought of as involving equivocations between high-level, functional objects or identity criteria, and low-level material objects or identity criteria. Distinguishing these suggests a levels-based solution to puzzles of material constitution. In the ‘different objects’ version, the solution is to relativise identity (and spatial exclusion) to levels of objects. On this suggestion, Theseus’s ship is identical with the aluminium ship in the functional sense, and with the museum ship in the material sense.

This suggestion can be carried over to all puzzles that involve different levels. It must be supplemented with an indeterminacy story for transitivity of identity puzzles that do not involve level differences. Intuitions about sameness of levels should determine which cases call for which kind of solution. If the suggested solution works, this lends some justification to the levels framework on which it is built.

Ch. 4: Wright and Johnston on response-dependence

After all this stage setting, it is time to return to our main topic. As mentioned in Ch. 1, response-dependence theses come in very different forms; the main divide is between theories like Wright's and Johnston's that imply that the very subject matter under discussion is somehow dependent on human responses, and theories like Pettit's that concern concepts only. In this chapter, we shall examine Wright's and Johnston's accounts of response-dependence. These are the two most influential and worked-out theories of response-dependence of subject matter in the literature. For each of the versions, we shall examine the intuitive content the distinction is intended to capture, the formal apparatus employed (conditions etc.), suggested applications, and the main problems raised by the proposal. Other views of this sort will be mentioned in passing. We shall begin with Wright, as he offers the most thorough discussion of the various conditions etc. that are shared by other authors, including Johnston. The next chapter contains an exposition of Pettit's view, and a discussion of an extension of it that suggests itself in the light of the levels framework.

In Ch. 6, we shall return to response-dependence of subject matter. With the results from Ch. 5 as backdrop, we shall discuss various ways of drawing the distinction, the relationship between them, what the core ideas are, and how to make them operational. In Chs. 7 and 8, we shall turn our attention to two specific problems that arise for response-dependence theses.

1. Wright: The Euthyphro Contrast¹

Crispin Wright's version of the response-dependence distinction is named after Plato's dialogue 'Euthyphro', in which Socrates and Euthyphro discuss the nature of the pious. Both agree that pious acts are all and only those approved by the gods, but they disagree about the order of determination. Euthyphro holds that certain acts are pious because the gods approve of them; the gods' approving responses *determine*

¹ Unless otherwise stated, the exposition is based on Wright (1992), appendix to Ch. 3, p. 108-139. This, in turn, is based on his unpublished manuscript 'Notes on Basic Equations'.

the facts about piety. Socrates thinks it is the other way around: the gods approve of certain acts because those acts are pious; their responses *detect* the facts about piety.

This disagreement illustrates the intuitive distinction that Wright aims to capture. Response-dependent concepts are concepts that have their extensions *determined* (at least partly) by judgements made in favourable conditions. For response-independent concepts, by contrast, the relevant judgements merely *detect* extensions that are determined independently of them.

The distinction can be explained in terms of the distinction between criterially and referentially governed concepts; where many (though not all) response-independent concepts have their extensions determined by the extension of a substantial property, response-dependent concepts are a type of criterially governed concepts, and apply to whatever competent judges would judge them to apply to. The point is not to track a substantial property, even if correlation with such a property cannot be ruled out.²

Though Wright's distinction is formulated as a thesis about concepts, it is naturally understood as a response-dependence thesis of subject matter. Response-dependent concepts are not in the business of referring to substantial properties. The extensions of the concepts, and of the corresponding deflationary properties, are determined by best judgements. In this sense, the subject matter itself is response-dependent.

Wright's formulation in terms of concepts may be motivated by the following considerations: First, it prevents misunderstandings based on disagreement about the nature of properties. People who believe only in substantial properties might object to talk about response-dependent *properties* because response-dependent concepts do not correspond to any substantial properties, and hence, in their view, response-dependent properties do not exist. This is no serious problem, as claims made in terms of response-dependent properties might easily be translated into terms such people can accept. But formulating the distinction in terms of concepts rather than properties takes away a potential for misunderstanding. It also prevents misunderstanding in another respect: talk about properties whose extensions are

² This way of explaining the distinction was conveyed in conversation, and is mostly implicit in Wright's written work. Any infelicitous formulations are probably my fault.

determined by responses in C-conditions might suggest that the responses are somehow causally involved in bringing about the properties, whereas of course the connection is purely conceptual. (The formulation in terms of concepts may give rise to misunderstandings as well; more on this later.)

In *Truth and Objectivity*, the Euthyphro contrast is presented as a part of a larger framework. It is impossible to do justice to this framework in this brief introduction, but a few words are needed to place the Euthyphro contrast in the right setting. Wright's project is to create a framework that allows for different degrees of realism in different areas. A core idea is that all it takes for a discourse to be truth-apt is that it complies with a set of platitudes, among which are the following:

...that to assert is to present as true; that any truth-apt content has a significant negation which is likewise truth-apt; that to be true is to correspond to the facts; that a statement may be justified without being true, and vice versa... (Wright 1992, p. 34)

If a discourse complies with these platitudes, the further question arises what specific features (the property or properties corresponding to) the truth predicate has in the domain in question. Four distinctions are offered as means to assess these further features. The Euthyphro contrast is one of them. Another is the possibility of evidence-transcendent truth in the area in question. If evidence-transcendent truth is a possibility, this allows for a higher degree of realism than if truth and evidence in the area are too closely related to allow for this possibility to make sense.³ The third is the 'cognitive command' distinction. Briefly, the question is whether faultless disagreement is possible in the area in question – more precisely, whether it is a priori that, vagueness aside, if two people disagree, then at least one of them is guilty of ignorance or error. The thought is that if the discourse is representational, then disagreements have to be explainable in terms of 'divergent input', either in background information or in the conditions of subject or circumstance, which make at least one of the disputants represent the subject matter in the wrong way.⁴ The fourth distinction concerns wide cosmological role, i.e. whether and to what extent

³ Wright (1992), e.g. p. 77-79.

⁴ Wright (1992), p. 91-93.

the disputed concepts might feature in explanations of other things than, or via, attitudinal states (beliefs etc.) in subjects regarding propositions of the disputed domain. For colours, for example, the question is whether colours might feature in explanations of other things than colour judgements, beliefs about colours etc., or explanations via such states. The wider the cosmological role, the more evidence for realism (and the less reason to believe that the facts in question are relative to our particular cognitive and perceptual makeup).⁵ This system of distinctions provides an account of truth which allows us to accept as truth-apt all domains that appear so on the surface, but which still allows us to recognise a variety of differences in the way statements of various kinds qualify for truth, and to allow for more and less objective subject matters.

If a concept is response-dependent in the Euthyphronic sense, this entails a moderate realism about the disputed domain. The semantic realism question about truth-aptitude receives a positive answer; statements about the domain are truth apt. If true, they are true in virtue of certain human response-patterns. The same goes for the question about epistemic access to the subject matter; in many (or core) cases, we get things right. Furthermore, there are indeed facts to get right or wrong, as the states of affairs posited by the discourse do indeed exist – there are colours, values, etc., even if these turn out to be response-dependent. Only realism question 3b): ‘Do the states of affairs etc. exist independently of us?’ receives a negative answer; response-dependent features are had in virtue of the way the cases would be classified by competent judges in favourable conditions. The concepts apply to whatever best judgements ascribe them to, and in virtue of their being so classified. (In terms of the levels framework, this amounts to a high level location, though not a location with a referent; more on this in Ch. 6.)

This sort of moderate realist position might be expected to be of use in areas that are *prima facie* truth apt, but intuitively closely linked to a human perspective – the combination that motivated the response-dependence distinction in the first place. Wright’s standard examples of response-dependence include humour and colours, but not moral values, the domain that provided the original motivation for the

⁵ Wright (1992), p. 196-99.

distinction in Johnston and others (see later). Other areas in which Wright finds use for the distinction are intentional states and meaning, though the distinction must be adjusted to suit the specific requirements of these domains.

1.1 The order of determination test

How do we determine whether a given class of disputed concepts are response-dependent in Wright's sense? Saying that the concepts are response-dependent if and only if they have their extensions (partly) determined by judgements in favourable conditions is usually not enough; if we're in doubt about their status regarding response-dependence, we will also be in doubt about the order of determination between best judgements and facts. To make the distinction operational (i.e. able to help decide about a domain we are in doubt about), we need to find other characteristics of response-dependent concepts by which they can be recognised as response-dependent.

For this purpose, Wright offers what he calls the 'order of determination test'. For a concept to be response-dependent, there has to be a true 'provisional equation' of the form

$$C \rightarrow (F(x) \leftrightarrow S \text{ would judge that } F(x))$$

that fulfils the following conditions:

- 1) the equation is true a priori
- 2) the C-conditions can be specified in a substantial, non-trivialising way
- 3) the C-conditions can be specified without presupposing the extensions of the disputed concepts as already given
- 4) there must be no better explanation of these conditions being met than response-dependence. In particular, if this can be explained by infallible tracking of underlying features, the concept is not response-dependent.

Wright does not take this characterisation to be the final word on the matter. As noted, he offers a different suggestion for intentional states and meaning where the above characterisation seems too demanding. This will be discussed in section 1.8. In the following, we shall go through the original suggestion step by step. At every

point, there will be issues worthy of discussion (many of which have been raised in the literature). I shall briefly mention some and provide references, but I will skip the detail of such discussions and focus on giving an overview of Wright's proposal.

1.2 Biconditionals

Wright's starting point is what he calls 'basic equations' – biconditionals of the sort usually used with response-dependence claims:⁶

- (1) X is F ($\text{/falls under the concept 'F'}$) \leftrightarrow in conditions C , subjects S would judge that X is F

However, basic equations are later replaced with 'provisional equations', where the C -conditions, designed to accommodate requirements on both subjects and circumstances, are moved outside the scope of the biconditional:⁷

- (2) $C \rightarrow (X \text{ is } F \text{ iff } S \text{ would judge that } X \text{ is } F)$

The purpose of this move is to avoid 'conditional fallacy' problems that allegedly afflict the basic equations. Such problems are discussed in detail in Ch. 8. Very briefly, the problem is that the C -conditions' becoming fulfilled might alter the situation with respect to F . For example, assume for the sake of the argument that colours are response-dependent, and say that X is a piece of white photo-sensitive paper in a photo lab. By hypothesis, X is white, but it would not look white in C -conditions, as bringing it into the daylight conditions normally favourable to colour assessment would change its colour. By moving the C -conditions 'outside the brackets', this problem would be avoided. If we consider an object that is already in C -conditions, any changes it might undergo as a result of coming into C -conditions will already have happened, and so the object would look white only if it actually is white.

⁶ Wright uses ' $P \leftrightarrow$ if CS , then S would judge that P ' (1992, p. 115). He might object to the ' X is F ' formulation, as it might suggest that best opinions determine facts in a causal sense, and might suggest to some that substantial properties are at stake. But as long as we steer clear of these misunderstandings, the ' X is F ' formulation should be innocent.

⁷ This move was inspired by a similar one made by Carnap in a very different context.

For brevity, I shall follow standard usage and sometimes call both basic and provisional equations biconditionals, though the latter are really biconditionals embedded in conditionals.

Wright chooses *judgements* ('x is red', 'x is good', etc.) as the relevant responses. The reason is that these are the only responses that are available in each and every domain where the response-dependence distinction might be of interest. For many domains, there are also characteristic phenomenal responses (nausea, red-experiences, etc.). But for some domains, there are no characteristic phenomenal responses. In others, the relevant phenomenal responses may be related to the concept in question in a way that would jeopardise the independence condition (see below).⁸ By choosing judgements as the relevant responses, we get a distinction that applies as widely as possible.

What sort of conditionals are we dealing with in the equations? Wright doesn't say. The natural reading is that they are all subjunctive conditionals and biconditionals. The question of the nature of the conditionals has received surprisingly little attention in the literature on response-dependence. Since I shall ultimately downplay the role of response-dependence biconditionals, I won't take on the task of spelling out what conditionals are needed, or what semantics is appropriate for them.

1.3 The substantiality condition

It takes more than a true provisional equation to make a concept response-dependent. If the C-conditions are specified as 'whatever conditions it takes to make subjects judge that x is F just when x is F', the biconditionals will be trivially true even for paradigm examples of response-independence (and analogously for S and R, the other place-holders used in response-dependence biconditionals). Simon Blackburn

⁸ There may be domains where the responses *would not occur* without a grasp of the concepts, or where the responses *would not count as relevant* unless informed by the concepts. Modal concepts may be an example of the former kind, as possession of modal concepts might be a precondition for the conceivability intuitions that guide us on modality. An example of the latter kind may be the (very diverse) phenomenal responses that may be relevant to morality.

has put the point nicely by pointing out that the following two claims are equally trivially true:

X is good \leftrightarrow X is such as to elicit desires from people under the ideal circumstances, i.e. those under which people desire good things

X is good \leftrightarrow X is such as to elicit astonishment from polar bears under the ideal circumstances, i.e. those under which polar bears are astonished at good things⁹

So C-conditions must be specified in a substantial, non-trivial way. This is traditionally done by giving a list of conditions specified in terms neutral to the subject matter – conditions that a competent language user will know, at least implicitly, in virtue of her mastery of the concept. An example from Wright is the following, admittedly incomplete, recipe for the concept blue:

‘...the surface must be in full view, and in good light, relatively stationary, and not too far away; and the subject must know which object is in question, must observe it attentively, must be possessed of normal visual equipment and be otherwise cognitively lucid, and must be competent with the concept *blue*. In addition, the subject must be free of doubt about the satisfaction of any of these conditions...’¹⁰

This specification might be disputed, refined etc., but for now, we shall bracket the details. In Ch. 7, we shall discuss some alternatives to this way of specifying C-conditions, and discuss the substantiality requirement further. Wright was the first to make this requirement explicit, but it is now endorsed by all authors on response-dependence.

1.4 Necessity and a priority

The truth of biconditionals with substantially specified C-conditions is still insufficient to settle the order of determination. This much is true even for response-independent matters such as shape; it seems plausible that e.g. x is pyramidal iff,

⁹ Blackburn 1993, p. 273.

given favourable epistemic conditions (which are probably as substantially specifiable as those for colours), subjects would judge *x* to be pyramidal. How to distinguish between the two types of cases?

One initially promising answer is that we might expect a *modal* difference between the biconditionals for response-dependent and response-independent concepts. If best opinions determine whether ‘F’ applies to *x*, then the corresponding biconditional should be necessary; we should expect the facts about ‘F’’s extension to follow best opinions across possible worlds.¹¹ If, on the other hand, these facts are merely tracked by judgements in certain (substantially specified) circumstances, then it is presumably a contingent fact that judgements in these conditions get things right. This suggests that response-dependent concepts can be recognised by the *necessity* of suitably specified biconditionals, while the corresponding biconditionals for response-independent concepts will be at best contingently true.

Unfortunately, this nice, simple picture is undermined by the possibility that rigidification in the specification of C (and S and R) might make the biconditionals necessary even for response-independent concepts (more on this in Ch. 6).

Wright’s response to this problem is to replace the necessity condition with an a priori condition. In Wright (1992), very little is said in justification of the idea that the biconditional must be a priori for response-dependent concepts, and a posteriori otherwise. All that is said is that if the necessity is produced by rigidification, it is not possible to know the proposition (including C-conditions and all) a priori. But if concepts have their extensions determined by best responses, it should be possible to discover this just by reflection on the concepts, and hence the biconditionals should be knowable a priori.

An a priori condition is accepted by all authors on response-dependence.¹² As we shall see in Chs. 5 and 6, the a priori condition raises a similar problem as necessity: it might be argued that it can hold for irrelevant reasons, as a by-product of

¹⁰ Wright (1989), p. 247.

¹¹ Throughout this work, I shall adopt the common if questionable strategy of talking about possible worlds without wanting to commit myself to any particular account of what their nature is and in which sense (if any) they exist.

¹² Again, except Miscevic 1998.

the way the concept's reference is fixed. See Ch. 6 for more discussion of both conditions.

1.5 The independence condition

The biconditionals characteristic of response-dependent concepts have a feature which, in other contexts, might seem highly suspicious: the concepts under discussion appear on both sides of the biconditionals. If the biconditionals were intended to provide definitions or reductive analyses of response-dependent concepts, they would be vulnerable to charges of circularity. If their purpose was to convey the meaning of the concept to someone who didn't know what 'red' meant, then the appearance of 'red' in the explanation would render it useless or at least unhelpful. However, the purpose of the biconditionals is the less ambitious one of reflecting the way the concepts get their extensions determined. So they do not immediately fall prey to a circularity charge. Still, there is a circularity-related issue to address. In Wright's words,

... a Euthyphronist who is content to implicate the distinctive concepts of a discourse in his formulation of the C-conditions, in such a way that satisfaction of the C-condition turns on details of the actual extension of those concepts, has to expect the challenge: 'Show that the way that you have implicated those concepts is consistent with your overall thesis, that their extension is, at least partially, constrained by best opinion'. (Wright 1992, p. 122.)

The worry is that if presuppositions about the extensions of the disputed concepts are needed to specify what opinions are best, this suggests a non-Euthyphronic order of determination. How can best opinions be what determines the extensions of the disputed concepts if the latter have to be settled to determine what opinions are best? Wright's initial response to this challenge¹³ was to add a third requirement for response-dependence: *the C-conditions must be specified in such a way that their satisfaction is logically independent of the details of the extension of concepts of the disputed class*. This 'independence condition' can be met either

¹³ In 'Notes on Basic Equations', recycled in Wright (1992).

- 1) by avoiding use of the concepts altogether, or
- 2) by employing them only in intentional contexts.

(In the latter case, we must presuppose that they have some extension or other, but need not presuppose anything about what that extension is).¹⁴

Wright does not claim that this response is the only way to meet the challenge. It is presented as a conservative line that is sufficient to avoid the problem, but may not be necessary. This attitude seems appropriate, as the independence condition is too strong to be met in some of the areas where response-dependence theses initially seem most attractive. One such area is moral concepts. Any plausible specification of the C-conditions for this area would presumably need to include a requirement to the effect that the subjects are ‘morally suitable’ or similar. So moral concepts will fail the independence condition as it stands.¹⁵ Another relevant area is intentional states, for which Wright suggests an adapted version of the response-dependence distinction (see section 1.8).

1.6 The extremal condition

The fourth and last of Wright’s conditions is the ‘extremal condition’. It says that the biconditional must be *primitively* a priori. That is, its a priori (and its fulfilling the other conditions) must not be explainable in terms of truth-makers other than best opinions. This condition is designed to rule out cases in which the biconditional is made true and fulfils the other three conditions because judgements in C-conditions *infallibly track* the extensions of the concepts, and not because the extensions are determined by judgements in C-conditions. Something of this sort would be the case on a Cartesian conception of mental states, e.g. pain: barring perturbing factors (concentration on a difficult task, neural disorders etc.), a subject will judge that she is in pain if and only if she is in fact in pain (and this is a priori), but on the Cartesian view, this is because we are exceedingly good at tracking the facts regarding our own pains, and has nothing to do with response-dependence.

¹⁴ Boghossian (1989, p. 546-7) and Johnston (1993, p. 124) have both challenged Wright on this point. See also Haukioja (2000). Wright offers a response to the former two in (1992), p. 132-135.

¹⁵ Wright raises this problem in his (1988).

Another way of putting the point is that a response-dependence thesis should be a last resort; if an alternative explanation of the truth-conferring state of affairs is available, then the concept does not count as response-dependent. Only if the a priori connection between responses and extensions is primitive and not otherwise explainable, it is a sign of response-dependence.¹⁶

1.7 The order of determination test

Together, the provisional equation and the four conditions constitute Wright's order of determination test. It is summed up neatly in 'The Conceivability of Naturalism':

A concept, F, is *Euthyphronic* if [...] a provisional equation can be written for it meeting each of the following four conditions:

- (i) The provisional equation is true a priori, as a matter of conceptual necessity.
- (ii) The conditions, C, are specified in specific, substantial terms
- (iii) The satisfaction of the C-conditions is a matter that is independent of the details of F's actual extension.
- (iv) The provisional equation is *primitively* a priori – it admits of no proof from ulterior premises concerning F of such a kind as to vindicate the idea that the C-conditions merely enable a subject to keep infallible track of an independently determined extension. (p. 425)

If a concept meets all these conditions, and hence is response-dependent, it is said to *fail* the order of determination test. (This way of speaking may seem confusing, but the idea is that a concept that meets these conditions fails of a sort of objectivity possessed by response-independent concepts and their subject matter.)

This recipe delivers the expected results in some domains: Colours presumably come out as response-dependent as expected. Humour, too, fails the test as expected.¹⁷ Shape comes out as response-independent, as it fails the independence condition (since we must presuppose constancy of shape across several observations in order to appreciate the whole shape of a three-dimensional figure).

¹⁶ This section is based mainly on Wright (2002) and discussion.

¹⁷ One might argue, though, that the correct theory for humour is not a response-dependence account, but a (stronger) form of relativism.

However, in some cases, we get unwelcome results. As mentioned, moral values do not qualify as response-dependent because of the problem with the independence condition, though this was an area in which we might have expected useful results from a response-dependence thesis. Another area that does not quite suit the current form of the distinction is intentional states. In this area, Wright argues that a response-dependence thesis might yield very interesting results. But this requires some modification of the first, standard version of the order of determination test. In the next section, I'll describe the modifications required, and the motivation for a response-dependence thesis of intentional states. I'll include a little more detail and discussion than I've done thus far.

1.8 Euthyphronism about intentional states

Why hold a response-dependence thesis of intentional states?¹⁸ According to Wright, such a thesis is attractive because it promises to accommodate two aspects of intentional states that seem hard to reconcile. The first is that intentional states share their first person epistemology with phenomenal states such as pains, colour experiences etc. It is characterised by the following features:

- *Authority*: in normal circumstances, a person's own judgements are the highest court of appeal with respect to her intentional states; unless there is positive evidence that something has gone wrong (e.g. inconsistency with behaviour or other self-ascriptions of intentional states), if she says she intends/hopes/believes that Φ , she should be taken to do so.
- *Transparency*: if a person is in intentional state I, she will normally be in a position to know this.
- *Groundlessness*: Beliefs about one's own intentional states are not usually inferred on the basis of other evidence; it does not make sense to ask for *reasons* for a subject's beliefs about her own intentional states.

The second aspect is that intentional states are answerable to certain constraints about future performance. In this, they resemble dispositional characteristics like

¹⁸ By intentional states, I mean propositional attitude states – believing, desiring, intending, wishing, fearing etc. *that P*. By using intention as an example, I don't mean to give this state a special role.

courage, modesty, endurance and irritability, and differ from phenomenal states. An ascription of a certain intentional state can be overridden if it conflicts with the subject's behaviour and other self-ascriptions of intentional states. For example, if someone claims to hope that Φ , but reacts with horror to indications that Φ might be about to materialise, so much the worse for her claim to hope that Φ .

It is easy to find a theory of intentional states that explains one of these two sets of features. The first person epistemology would be nicely explained by a Cartesian view, on which the subject has direct, infallible access to her own mental states in virtue of a sort of observational privilege. But such a view fails to do justice to the conditionality on future performance. On the other hand, a dispositional account of intentional states would nicely explain the latter, but it would make a mystery of the first person epistemology of intentional states. The challenge is to find an account that would accommodate both groups of features.

A Euthyphronic account of intentional states seems well suited for this task. Authority is explained by the right to left reading of the biconditional in the relevant provisional equation: Given favourable conditions, if a subject would judge that she is in Φ , she is in Φ . Transparency follows from the left to right reading: Given favourable conditions, if she is in Φ , she would judge this to be so. Groundlessness is also accommodated, as subjects are not in the business of tracking independent facts about their intentional states; there is nothing more to e.g. having an intention than that given suitable conditions, we would judge this to be so. (Relatedly, you need no reasons for your self-ascriptions of intentional states; you are justified in making any judgement about them whenever it seems right to you because of the way the C-conditions work in this domain; more on this below.) The disposition-like features of intentional states, on the other hand, are accommodated by the possibility that conditions might turn out not to be favourable after all. Such a conclusion can be forced by inconsistency with the subject's behaviour or her other self-ascriptions of intentional states.

The core idea of a Euthyphronist thesis of intentional states is the usual one: best opinions determine the extensions of the concepts, rather than tracking independently constituted facts about the subject matter in question. But the exact shape of the proposal must be slightly different from the general Euthyphronist

paradigm. The suggestion can't be that judgements about one's own intentional states made in circumstances that can be checked in advance to be favourable determine extensions case by case. As pointed out by Boghossian (1989), the alleged extension-determining judgements are themselves states of exactly the kind under discussion. These judgements must have a specific content if they are to be able to do any extension-determining at all. But doesn't this mean, contrary to Euthyphronism, that the extensions of (at least some) concepts of intentional states, and the corresponding facts about intentional states, must be given in advance of these judgements being made?¹⁹

Wright's response to Boghossian's worry is that ascription of content to intentional states is a holistic matter. Unfortunately, his written comments on this point are very brief. He writes:

My own instinct is that we do better, in the case of intentional states, to look for a holistic mode of dependence: roughly, that the details of a subject's intentional states are, a priori, determined in such a way as to maximise harmony with her self-conception, as manifest in her own elicitable self-ascriptions (or, at least, to minimise inexplicable discord with it).²⁰

The idea seems to be that the contents of intentional states are not to be determined case by case by the deliverances of best opinions in specific circumstances. Rather, the intentional states of subjects are in principle assessed as a whole, which allows the various states to have a role in determining each other.²¹ The subject's self-ascriptions of intentional states should be taken at face value unless something contradicts them, in which case the revision should be chosen that allows her as much rationality as possible.²²

¹⁹ Boghossian (1989), p. 546-47.

²⁰ Wright (1992), p. 138-39, n. 47.

²¹ The Euthyphronist account of intentional states thus has affinities with Davidsonian interpretativism: someone is in a certain intentional state is that the most charitable overall interpretation of the subject's mental life entails that she is in that state. Wright's view differs from Davidson's in that the first person point of view is given a constitutive role in the account.

²² An alternative understanding of the account would have it that C-conditions for judging on a subject's mental states include being that subject; this would be a way of cashing out the privileged position of the first person. Wright has made it clear in discussion that this is not what he intends,

The same line of thought underlies Wright's strategy with another problem for Euthyphronism regarding intentional states: with intentional states, there is no prospect for a complete set of substantially specified C-conditions. To capture the class of best judgements about intentional states, we would need to rule out perturbing factors like self-deception. But self-deception functions as a catch-all clause for all the ways in which a subject could be ignorant or mistaken about her own intentional states, and so is something very close to the whatever-it-takes-clauses that the substantiality requirement was meant to rule out.

Wright's response to this problem is that the no-self-deception condition is positive-presumptive: if there is no evidence to the contrary, we assume that it is fulfilled, whereas it takes evidence to show that it is *not* fulfilled. Similarly for other parts of the C-conditions, e.g. that the subject must be paying attention to the relevant states. This means that in the absence of evidence to the contrary, we can assume a biconditional connection between intentional states and the person's own beliefs about them. In effect, the no-self-deception clause and similar conditions can be deleted from the list of C-conditions.

This move fits well with the way discourse about intentional states works: When assessing self-ascriptions of intentional states, we actually do take them at face value as long as nothing counts against them, though we are willing to revise them in the face of countervailing evidence. However, there is a question whether this move makes the C-conditions work in a too flexible way; in effect, it enables us to claim failure of C-conditions every time the biconditional is violated. Is this consistent with the motivation behind the substantiality condition?

Probably the answer is yes. As we shall see in Ch. 7, a too rigid interpretation of the substantiality condition is both unnecessary and too strong for many cases where response-dependence theses seem attractive. The condition is there merely to rule out triviality, and this is something far weaker than requiring a full list of conditions that can be known independently to obtain.

If the move with positive-presumptive C-conditions works, it might also address another challenge to Euthyphronism about intentional states, this time

however. His interest in the first person perspective notwithstanding, he wants to reserve a more important role for interpretation from a third person's perspective than this suggestion would allow.

regarding the independence condition: It is not possible to specify the C-conditions in a way that is independent of the extensions of concepts of intentional states. For example, it is hard to see how to capture the notion of self-deception without using intentional state concepts in illegitimate ways.²³

The move with positive-presumptive C-conditions promises a way out. If each of the problematic C-conditions can be assumed to be met if nothing suggests otherwise, and if this allows us to delete it from the ‘pool of assumptions’ in the C-conditions, then we can presumably get the concept off the ground without making illegitimate assumptions about its extension. When later (logically and temporally) we may have to judge that the conditions were not in fact met, enough of the content of the concepts will already be in place, and so any presuppositions about it will not be illegitimate. Or at least that would be the suggestion.²⁴

Euthyphronist accounts of intentional states deserve further attention, and raise further problems; all I hope to have done in this section is to convey a flavour of the proposal, and the way it differs from the standard Euthyphronist recipe.

1.9 Euthyphronism about meaning

Wright’s Euthyphronist proposal for intentional states, with its adjustments to the standard Euthyphronist recipe, can be used as a template for a Euthyphronist account of meaning. As with intentional states, Wright does not himself endorse such an account, but merely tables it as an interesting but not unproblematic possibility that deserves further investigation.

²³ Alex Miller (1989) has voiced similar objections, and offers a more elaborate discussion of the problem. If the move with positive-presumptive C-conditions works and is applicable to all the problematic C-conditions, it would presumably address Miller’s worries as well as this simplified version.

²⁴ The move with positive-presumptive conditions may seem unsatisfactory if viewed as an answer to the isolated problems of meeting the substantiality and independence conditions as stated. However, the interesting issue is not whether the letter of these conditions can be respected, but how the substance of the proposal – the intuitive distinction Wright aims to capture, and the intuitive considerations that motivate the conditions of the order of determination test – can be applied to the case in question. This should be the main focus in a further investigation of the prospects for an Euthyphronist account of intentional states. I won’t take a view on whether it works.

Meaning something by a linguistic expression presumably is, or involves as a component, an intentional state. Hence, to the extent that concepts of intentional states are response-dependent, we might expect response-dependence for the concept of meaning (in this sense) as a corollary. However, there are other reasons to find such an account interesting. The main motivation for the proposal (and for Pettit's similar one; see Ch. 5) is that it might provide a solution to the Wittgensteinian problem of rule-following.

The problem can be summed up as follows: On the Platonist conception of rules targeted by Wittgenstein's discussion, a rule (linguistic or otherwise) determines the correct application to any new case independently of what anybody might think on the matter. It can be pictured as a set of rails stretching into infinity; getting the rule right is to stay on a track laid out independently of any applications.²⁵ But on such a conception, it becomes a mystery how language users could latch on to the rule, or how they can know its requirements in new cases. (After all, any finite set of examples used in learning a rule is compatible with infinitely many rules, and any rule is compatible with any course of action under some interpretation.)

In response to this problem, a number of alternative paths have been explored. One (extreme) reaction is irrealism regarding rules – the view that nothing makes one particular course of action more correct than any others. This can be combined with a 'sceptical solution' à la Kripke, describing how the practice of ascribing meanings can still play an important role in our lives. Another option is consensualism, the view that rules are constituted by consensus in a community. These views would avoid the epistemological problem with accessing the requirements of the rule, but at the price of sacrificing its objectivity. A third option is quietism – rejection of constitutive questions altogether. None of these positions seem satisfactory.

An Euthyphronist account of rule-following might provide a reasonable compromise between intuitions about the reality or objectivity of rules and intuitions about epistemic access (our ability to read off their requirements). On this view, rules will have some sort of objectivity in that not anything goes, and there is a distinction between what seems right to a subject in a given case and what is right. Epistemic

²⁵ Wittgenstein (1953), §218; Wright (2001).

access to the requirements of the rule is also guaranteed to a suitable degree, as in C-conditions, we get things right.²⁶

Developing a satisfactory Euthyphronic account of meaning would take some work, however. First, there is the question what judgements are relevant. Is it a person's judgements of own meanings? Or third person judgements about what someone means by an expression? Judgements about what an expression means in itself? In order to make the proposal work, this should be made clearer.²⁷ A further complication is that for many rules, it may not be possible to isolate a well-defined class of characteristic judgements that do the extension-determining; there may be no such characteristic judgements at all. This may not be fatal to the proposal; in the case of rule-following, we should expect the same kind of holism as in the case of intentional states. But the issue would need further attention.

Other challenges would arise regarding the 'formal' requirements for response-dependence. First, there would be a problem with substantiality exactly analogous to the one discussed for intentional states. Secondly, there would be a problem with the independence condition: In order for a specification of the C-conditions to say anything at all, we must presuppose the extensions of the concepts used in that specification. But that means presupposing facts about meaning as already given. In both cases, a solution might be sought along the lines described above: holism and positive-presumptive conditions. For the independence problem, the idea would be that if all the C-conditions in the case of meaning are of the positive-presumptive sort, they might all in effect be deleted, and hence no concepts are needed to state them. The same reservations would apply as for intentional states: the important issue is not respecting the letter of the conditions, but whether the overall proposal can be made to work.

A further, and seemingly serious, problem for Euthyphronism about meaning is that it threatens to globalise: it might be argued that if meaning is response-dependent, all concepts are response-dependent. This would be an unwelcome

²⁶ Wright has said little directly about the motivation for a response-dependence account of meaning; this motivation is my reconstruction.

²⁷ This problem was stated by Crispin Wright in discussion.

conclusion. First, it would undermine the motivation for Wright's project of providing a system of distinctions that allows us to distinguish between different degrees of realism in different domains. The distinction is useless for this purpose if all cases fall on the same side. Secondly, global Euthyphronism would be a strongly anti-realist conclusion. (In this, it differs from Pettit's response-dependence account of meaning, which is intended to apply globally and does not seriously compromise realism. So if Pettit's account can do the job, it is presumably preferable over a Euthyphronist account.)

There are various arguments that response-dependence of meaning must globalise. Wright presents and discusses the following simple and intuitive version:²⁸ The truth value of a sentence is normally thought to be a function of two components: The meaning of the sentence and the state of the world in the relevant respects. It is natural to think that if there is some sort of limitation on the realism level for the first component, then that limitation will carry over to the whole, i.e. to the truth of the sentence. If, for example, the rule-following considerations force a non-factualist conclusion regarding meaning, then presumably non-factualism about truth follows, and hence non-factualism for any proposition about any subject matter. The same line of thought might be applied to other limitations, such as response-dependence: If the meaning of sentences about a given domain depends on best opinions, the latter will also play a role in determining the extension of the truth predicate among those sentences. Or so the argument goes.

This line of thought can be, and has been disputed.²⁹ However, a fuller treatment of the issue is beyond the scope of this chapter. We shall proceed to look at some other general objections to Wright's view.

²⁸ Wright (1984) and (1992). The 1984 version (p. 769) focuses on non-factualism about meaning. A general version of the argument is discussed in (1992), p. 212. Boghossian (1989, p. 524-25) has given a more detailed argument for the same conclusion. For a criticism of his argument, see Wright (1992, p. 214-20).

²⁹ Wright goes on to argue it is not always *limitations* on realism that are infectious. For the cognitive command constraint, limitations are infectious, but not so for evidence-transcendence. And he sees no reason to suppose that Euthyphronism is infectious; just because there is a suitable provisional equations that fulfils the conditions for P, there doesn't need to be one for a compound including P, e.g. P & Q. I suspect that how the various distinctions fare in this respect may well be an

1.10 Some objections

In this section, I shall provide a brief overview of some problems raised by Wright's version of response-dependence, apart from those already mentioned. Some of them will be discussed later in the thesis.

The univocity objection

Wright's provisional equations say, not that extensions are determined by judgements in C-conditions, but that in C-conditions, judgements determine extensions. This raises the question what determines whether the concept applies in cases where the C-conditions are not met. And if something different does the extension-determining job for these cases, with what right do we say that there is one, not two concepts in play? The natural response is that 'the concept applies if it would be judged to apply if the C-conditions were met'. But this answer would raise conditional fallacy problems. Accordingly, Wright has shifted the focus to provisional equations, and thus barred himself from giving this answer. How, then, can the challenge be met? We shall discuss this problem further in Ch. 8.³⁰

Reconciling substantiality with a priority and independence

A second problem is a prima facie conflict between the substantiality requirement and the a priority requirement.³¹ The gist of the problem is this: In order to be anywhere near complete, a list of substantial C-conditions, on the recipe of Wright's suggestion for colours above, would need to contain material that we cannot know a

artefact of the way they are formulated – in particular, a function of whether a set of sufficient conditions are given for qualifying for the *higher* degree of realism (which is the case for cognitive command) or a set of sufficient conditions for realism being *limited* in a certain way (as is the case with the Euthyphro contrast).

There are also other features that need to be taken into account. In the case of Euthyphronism, it makes a difference what exact group of judgements and C-conditions we are talking about. Dependence on best judgements about meanings is not the same as dependence on best judgements about another subject matter, e.g. colours, humour, or whatever. So even if there is a sort of response-dependence that is infectious, it is probably not response-dependence in the sense of failing the order of determination test.

³⁰ The problem was raised by Wright (1992), p. 125.

³¹ This problem is raised in Pettit (1991), p. 603.

priori. For example, it is presumably not a priori that a black and white disc would seem to change colour when spinning fast, and so that such cases need to be excluded by the C-conditions. How can we take such cases into account in a way that respects both the substantiality and the a priori constraint? This problem provides the stage setting for Ch. 7 on C-conditions. It poses a challenge to other response-dependence accounts as well as Wright's. A similar problem arises about reconciling substantiality and independence.

The content of extension-determining judgements

A third problem, also related to independence considerations, has been raised by Mark Johnston. The worry is similar to Boghossian's objection to Euthyphronism about intentional states. But Johnston's worry targets Euthyphronist theses in general. The worry is this: The disputed concepts feature in the alleged extension-determining judgements. If these judgements are to be able to determine anything at all, the concepts they involve, including the ones under discussion, must already have content and extensions. And, the thought is, this extension cannot be determined by best judgements, because it is a prerequisite for these judgements having any content at all.³²

Wright's response to this objection is that it gains purchase if, but only if, the order in question is a temporal one. If the order is temporal, it is correct that the concepts involved in the judgements must have their extension fixed before they can determine anything else. But if the order is logical, the situation is different. There is still a constraint like the one that motivated the independence condition: the concepts must not be used in ways that are inconsistent with a Euthyphronic order of determination. But this constraint can be met in less demanding ways than by prohibiting the use of the concepts altogether. For example, as Wright suggests, it could be done by allowing them only in intentional contexts. The thought seems to be that such use would presuppose that the concepts have *some* extension, but not *which* extension they have, or how that extension is determined. Perhaps this suggestion could be supported by a holistic story like the one invoked in the reply to

³² Johnston (1993), appendix 3, p.124.

Boghossian's objection. This might amount to taking on board Johnston's point³³ that the relation between best judgements and facts must be *interdependence* rather than one-way dependence, at least in the case of intentional states. But Wright could presumably do this without changing the spirit of his proposal.

Apart from the problems mentioned here, there are also a couple of more general objections to the idea of response-dependence of subject matter. One is the conditional fallacy problems which are the subject of Ch. 8. Another is Johnston's missing explanation argument. This will be presented in the exposition of Johnston's view, to which we shall now turn. The exposition of Johnston's account will be briefer than that of Wright's because Johnston's core ideas are simpler to explain than Wright's, and because some relevant issues, e.g. the motivation for the substantiality, a priority and necessity conditions, have already been discussed.

2. Johnston: Response-dispositional concepts

Johnston's account of response-dependence has many affinities with Wright's, but is also importantly different. As mentioned in Ch. 1, Johnston originally tried to capture the distinction in terms of a contrast in the direction of reading of basic equations, which Wright developed into the order of determination test.³⁴ The two accounts thus have a common origin. Johnston later gave up on the 'direction of reading' idea, and has expressed scepticism about Wright's order of determination terminology because of the worry outlined above.

Johnston's version of response-dependence is the one that comes closest to the traditional Lockean distinction between primary and secondary qualities. It was motivated by the project of finding an account of moral values that reconciled their subject-relatedness with (a moderate) realist stance towards them. Johnston (1989) continues McDowell's and Wiggins' project of seeking such an account by an analogy with secondary qualities.

³³ Johnston (1993), p. 126.

³⁴ Wright (1992), p. 109.

On Johnston's view, response-dependent concepts are *concepts of response-dispositions*. In (1993), he formulates his view in terms of concept identity claims. A concept is response-dispositional iff

The concept 'F' = the concept of the disposition to produce R in S under C, where

- 1) the manifestation of the disposition, R, is 'some response essentially and intrinsically involving some mental response' (i.e. not purely bodily responses like sweating and digesting),
- 2) S is some subject or group of subjects,
- 3) C is some specified set of conditions, and
- 4) the identity is not due to trivialising 'whatever-it-takes' specifications of R, S and C.³⁵

The specifications of S and C can be rigid or flexible as it suits the case in question, and can pick out statistically typical or idealised conditions or observers. The responses can be judgements or perceptual responses. Depending on the choices made on these matters, the resulting response-dependence claims can be very different. Truth-functional or quantificational compounds involving response-dispositional concepts are to count as response-dependent in a broader sense. Concepts that do not involve response-dispositions in any of these ways are response-independent.³⁶

What is the status of these conceptual identity claims? ³⁷ As Johnston points out, they are hopeless as analyses or reductive definitions, since the concept under

³⁵ Johnston (1993), p. 103.

³⁶ Johnston (1993), p. 103-4. In the context of our discussion of globalisation arguments above, this means that response-dependence, not response-independence, is contagious.

³⁷ What exactly do such concept identity claims say? 1) That it is a priori that two concepts – one of the disputed class, and one that is explicitly a concept of a response-disposition – refer to the same property? Or 2) that the two concepts share the same sense, i.e. that if they are criterial, they pick out the same instances as falling under them, and if referential, they pick out the same referent in the same way? Or do they say 3) that there is an a priori connection between a concept (the disputed

discussion is overtly or covertly presupposed on the right hand side. This has been taken by some to imply a questionable and possibly vicious circularity. However, the biconditionals were never intended as analyses. In Johnston's version, their purpose is just to state an interdependence between the concept under discussion and the concept of a disposition of ours to respond in certain ways. For this purpose, the 'circularity' involved is innocent. Johnston holds that '...circularity of the sort indicated would be a defect only if it made the biconditionals and their associated identities empty. But this is evidently not the case. The biconditionals are very strong claims.'³⁸

Response-dependent concepts being concepts of response-dispositions means that response-dependent concepts refer to response-dependent properties. In (1998), Johnston makes this explicit, and changes the focus to response-dependence for properties, not for concepts. The view is that

...a property, Being F, is response-dependent if there is some predicate 'is f' which expresses the property (i.e., whose extension across possible worlds is just the things which have the property) such that a substantial way of filling out 'R', 'S' and 'C' makes

x is f if and only if x is disposed to produce x-directed response
R in all actual and possible subjects S under conditions C

a priori and necessary. (Johnston 1998, p. 9.)

(Note that the 'would' used in Wright's version of the biconditional has been replaced with 'is disposed to'. The purpose is to avoid conditional fallacy problems.)

Johnston shares with Wright the a priority and substantiality conditions, and adds a necessity condition. The rationale of the latter is that in all worlds, the concept will apply to just the things that would elicit the right responses from the right subjects in the right conditions, as the concept is a concept of a disposition to elicit

one) and its referent (the response-disposition)? On the whole, 3) seems the choice that makes most sense of Johnston's views. These and related issues will be discussed briefly in Ch. 6.

³⁸ Johnston (1993), p. 106. A similar defence might be used by Wright and others (with suitably different 'job descriptions' for the biconditionals).

those responses.³⁹ Johnston adds an explicit requirement that the necessity must not merely be a result of rigidification.⁴⁰ Nothing in this new way of putting matters seems incompatible with the earlier account.

2.1 Descriptive and revisionary response-dependence theses

Johnston describes two ways response-dependence theses can come about. Some concepts, such as ‘nauseating’, wear their response-dependence on their sleeves; it is clear to all competent users of the concept that to be nauseating is to have a disposition to elicit nausea in people. A response-dependence thesis about such a concept is a *descriptive* thesis about how the concept actually works. Johnston names this type of view ‘descriptive protogoreanism’ after Protagoras’ alleged claim that man is the measure of all things – a claim that may well have been intended as a statement of a response-dependence thesis or something close.

‘Nauseating’ is not, strictly speaking, a response-dispositional concept in the intended sense; for this, the responses must be cognitive rather than bodily ones. But as Johnston notes, it is hard to find an uncontroversial example of a cognitively response-dispositional concept. Johnston holds that few if any concepts are (cognitively) response-dispositional as they are commonly used. He supports this view by considerations about how the candidate concepts work. For one group of candidate concepts – referential concepts that work on the paradigm of natural kind concepts – the problem is exactly that the concepts are referential, i.e. aim to pick out a property which then determines the extension. This property may perhaps be picked out as the one normally responsible for responses in C-conditions. But the property itself is not a response-disposition.

For another group of candidate concepts – cluster concepts whose content is given by a set of core intuitions, without which we would lose our grip on the subject

³⁹ This view would not sit well with a thesis that dispositions are identical with their bases. More on this in Ch. 6.

⁴⁰ Johnston (1998), p. 10. Unless a story is told about how we can tell whether the necessity in question is merely an artefact of rigidification, or whether it has come about in a more interesting way, this requirement comes across as a wish list rather than a tool for making the distinction operational. Wright and Wedgewood face a similar challenge; more on this in Ch. 6.

matter in question – the problem is that the core intuitions will often include (combinations of) intuitions that rule out response-dependence, e.g. intuitions about causal efficacy.⁴¹

However, for some concepts, it might turn out that no property satisfies all intuitive demands. In such cases, the most charitable alternative is often to replace the *prima facie* response-independent concept with a response-dispositional counterpart. This might happen e.g. when the platitudes essential to the concept imply that the concept refers to a simple, unified physical property of objects, and it turns out that in reality there is no appropriate common factor on a physical level. In such cases, a *revisionary* response-dependence thesis will be motivated: a thesis that the original (response-independent) concepts are faulty and should be replaced with response-dispositional counterparts. In ‘How to speak of the colours’ (1992), Johnston argues that this is the case for colour concepts. As they stand, colour concepts fail of reference, since the core intuitions associated with them can’t all be satisfied together. (In particular, according to Johnston, a core intuition is that colour experience reveals all there is to know about colours, and this is at odds with all possible candidates for a colour theory.) So we should choose the most charitable revision of the concepts. This turns out to be a response-dispositionalist view. Though this view entails minor intuitive compromises (such as contradicting the intuition that colours are intrinsic properties of objects, not relational properties involving subjects), response-dispositions are the best candidates for the reference of colour concepts, as their physicalist rivals would compromise intuitions even more.

In our levels terminology from Ch. 2, we can reformulate the story as follows: the colour concepts fail of reference because they come with core intuitions that point to several different levels, and thus lead to contradictions when taken together. The best way to resolve this conflict is not to eliminate the colour concepts, but to endorse a revisionary response-dependence thesis, explain away a few intuitions pointing to different locations, and conclude that colour concepts are concepts of response-dispositions.

⁴¹ Johnston (1993), p.107.

Johnston's response-dependence thesis of values comes about in a similar way. The thesis is that 'x is a value iff substantive reason is on the side of valuing x', where substantive reasons are given a response-dependence account. Something is a substantive reason iff we are disposed stably to take it to be so under conditions of increasing information and critical reflection, and a similar response-dependence account is given for acceptable methods for determining whether the weight of substantial reasoning is on the side of valuing x.⁴² We might have expected something more objectivist than this, but this turns out not to be feasible.

2.2 Response-dependence and pragmatism

In 'Objectivity Refigured: Pragmatism Without Verificationism', Johnston argues for another wide-ranging application of response-dependence: a response-dependence thesis of '*accessible truth*'. Truth itself is response-independent according to Johnston, and he attacks verificationism, internal realism and similar theories that can be construed as response-dependence theses about truth (i.e. global response-dependence theses). However, he wants to retain the pragmatist insight that all the truths we should care about are those that are in principle accessible to us: those we will arrive at (i.e. are disposed to accept) at the end of rational enquiry. Unlike truth simpliciter, this notion of accessible truth is a response-dispositional notion. Another pragmatist point that he wants to retain is what he calls The Metaphor of Nature's Toleration – in his own words, 'the idea that Nature does not *demand* to be theorised about or described in one specific way, so that the *rightness* of employing a theory and hence the scheme of description it embodies is not the solitary work of nature but is in some sense up to us'.⁴³ Rightness, even in epistemic matters, is a normative notion. As a corollary of the response-dependence account of value, we get a response-dependence thesis for the rightness of accepting a theory:

The act of accepting for some given purpose and in some given situation a theory T and the scheme of description T embodies is right iff we are stably

⁴² Johnston (1989), especially p. 162-64.

⁴³ Johnston (1993), p. 113; his italics.

disposed to find this act right under conditions of increasing non-evaluative information and critical reflection.⁴⁴

This is a surprising consequence of a response-dependence theory of value. Normally, a response-dependence theory of value would be taken to be a fairly local conclusion, relevant only to ethics, maybe aesthetics, and possibly a few other domains. But Johnston takes it to include epistemic value, and hence all the truths we should be interested in – as opposed to those we shouldn't concern ourselves with, e.g. because they are in principle inaccessible, or because they are 'horrible, depressing, or despicable thing[s] to believe or even to entertain'.⁴⁵

This view seems highly controversial. First, truth and accessible truth in Johnston's sense – super assertibility, in Wright's terms – might presumably come apart, and not just by superassertible propositions being merely a subset of true ones. There could presumably be sentences that are superassertible but false. Where does that leave e.g. the platitude that we should aim at forming true beliefs?

A second, related objection is brought up by Johnston. It is that what makes it right to accept a certain theory is not something about responses, but simply that the theory is true. Epistemic value does not seem to depend on response-patterns. Johnston's (slightly disappointing) answer is that the value of accepting a theory depends on our interests, and that our interest is rarely in discovering the truth at whatever cost. Even where this is our interest, what makes it right to accept theories because they are true is the interest we have in the truth.

A better answer might be that what makes it right to accept a theory is not its truth in itself, but our reasons to think it is true. Epistemic virtue, seen from the inside, works a bit like other normative domains. It is not about getting it right, by pure guesswork if need be, but about using methods that usually work and are known as the most reliable means of getting as much as possible right. So it is not truth in itself that makes a belief a good one; it is our reasons to think it is true. This line of

⁴⁴ Johnston (1993), p. 114. This formulation is later (p. 115) given the proviso 'Fiendishly pure theoretical purposes aside', to bracket situations where our interest is only in knowing the truth, whatever the cost to our interest. Johnston argues that such an interest is rarely justified, and that whether it is justified is itself a response-dependent matter.

⁴⁵ Johnston (1993), p. 115.

thought – ‘truth in itself is best aimed at by aiming at superassertibility’ – may also provide an answer to the first worry. It is not clear that these answers are satisfactory; there is a long discussion to be had over the mileage of Johnston’s controversial suggestion. But this is not the place for that discussion.

2.3 The Missing Explanation Argument

Important as they are, these response-dependence theses are exceptions. Johnston does not believe in widespread or global response-dependence. In his view, ‘only if fundamental reality were intrinsically mental could a set of concepts adequate for its description consist only of response-dependent concepts.’⁴⁶ This seems too strong, not the least because ‘response-dependent’ in Johnston’s vocabulary includes non-response-dispositional concepts with response-dispositional components. But it is true that a global Johnstonian response-dispositionalist thesis would entail an unpleasant form of anti-realism; it would entail that all the properties we have concepts of are response-dispositions (and if the world contains properties apart from these, our concepts are not ‘adequate for its description’).

A main reason for Johnston’s scepticism about widespread response-dependence is the ‘Missing Explanation Argument’ (MEA in the following). The core idea behind the argument is that a priori equivalents cannot also be related as explanandum and explanans in empirical explanations. For example, if it is a priori that something is red iff it is disposed to look red to standard subjects in standard conditions, it cannot also be the case that an object is disposed to look red to standard subjects in standard conditions *because* it is red. For each domain, we have to choose: either the causal claim or the response-dependence thesis has to go.

Johnston has presented the argument in two different versions. The first can be summed up as follows:⁴⁷ A response-dependence thesis of F’s claims a priori equivalence between

1) x is F

and (something like)

⁴⁶ (1993), p. 106.

⁴⁷ Johnston (1991), p. 130-31.

2) x is disposed to seem F to S in C

In many domains where response-dependence theses seem attractive, it is a deeply entrenched intuition that there are true empirical explanations of the form

3) x seems F (to S in C) *because* it is F

A further premise is the following substitution principle:

4) substitution of a priori equivalents into true empirical explanations may make them false, but cannot turn them into ‘explanatory solecisms’ – claims that are a priori hopeless as empirical explanations.

However, if we substitute 2) for 1) in 3), an explanatory solecism is exactly what we get. So one of the claims must go: the response-dependence thesis or the empirical explanation.

Johnston applies this argument in many domains where such empirical explanations are intuitively indispensable, and hence (in Johnston’s view,) response-dependence theses won’t work. The most wide-ranging application is against internal realism and similar response-dependence theories about truth.⁴⁸ In some cases, though, such as colour and value, the axe falls on the empirical explanation rather than the a priori equivalence.

The missing explanation argument has affinities with the causal exclusion arguments described in Ch. 2. But it differs from these in that its scope is narrower: First, the argument is directed against one specific location option, RD theses, and not against high level locations in general, whereas causal exclusion arguments can be applied at any level but the lowest (if a lowest level exists). Secondly, it is not claimed that response-dispositional properties are unable to cause anything at all, but only that they are unable to cause a particular kind of facts: the characteristic responses. (This is compatible with the view that, qua dispositions, they can’t cause anything at all, but it does not force such a conclusion.)

The first version of the MEA has been heavily criticised. Some critics have focussed on the substitution principle. Miller (1995) has argued that while perhaps

⁴⁸ Johnston (1993), p. 94-95, and (1991), p. 173-78.

usable in some contexts, the substitution principle will not work for the cases it is intended for. McFarland (1999) has argued that the principle fails for independent reasons. Other influential criticisms include Pettit & Menzies (1993), who argue that the argument hinges on an equivocation (understanding F as *having* a disposition in some places and as *manifesting* a disposition in others), and that the result of the substitution, consistently applied, is a true though uninteresting empirical explanation: x looks red (i.e. manifests the disposition of looking red) to S in C because it has the disposition to look red to S in C; this is exactly as we should expect, given the equally uninteresting explanation that was the starting point.

In response to these and other criticisms,⁴⁹ Johnston has presented a new version of the argument (Johnston 1998) which does not involve the substitution principle. I will skip the detail of this version, since it is more complex than the first version, but seems to do no better in proving the point.⁵⁰

Despite the shortcomings of Johnston's precise formulations of the MEA, there seems to be an important point in the vicinity which is brought into focus by the levels framework: if a property F is response-dependent, then it is not, strictly speaking, F that does the causal work in bringing the characteristic responses about. The causal work will be done by lower level properties on which F may supervene (in some appropriate sense), but which are not identical with F. There is, of course, a weak sense in which it can be true that something appears, say, red because it is red: the sense for which the alternatives are 'x appears red because the light is weird, because you're wearing those funny glasses, etc.' But if we're talking causality proper, the causal work is not done by the response-dependent property, but by low level correlates. In Johnston's version, the point is the familiar one that dispositions don't cause their manifestations in any substantial sense; the causal work is done by

⁴⁹ For other criticisms, see Blackburn (1993), section 3 (p. 267-271); Haukioja (2000), Ch. II.2.2, p. 106-112; Wright (1992) p. 128-132; Pettit (1996), Lopez de Sa (2003), Ch. II.5, p. 70-73. For replies to some criticisms, see Johnston (1998), p. 24-35.

⁵⁰ For the argument, see Johnston (1998), p. 15-23. For criticisms, see Haukioja (2006), who argues that the explanation that goes missing in this version of the argument is not in fact an empirical one, and Miller (2001), who argues that the argument from Menzies & Pettit (1993) can be adapted to work against the second version too. See also Lopez de Sa (2003), p. 76-84.

the causal bases of the dispositions (again, this won't be true if dispositions are identified with their bases, but this view won't serve Johnston's purposes in any case). In the case of Wright-style response-dependence, the point is even more obvious. The deflationary properties correlated with criterially governed concepts are not the right kind of properties to be causes. Something qualifies as F in virtue of the fact that it would elicit the appropriate responses in C-conditions. But how they are caused is a different matter, and the causing is probably best thought of as being done by low-level, substantial properties.

Over the years, the sceptical line based on the 'missing explanation' seems to have grown stronger in Johnston's work. In the paper that presents the second version of the missing explanation argument, he calls response-dependence a 'sophisticated twist on projectivism'.⁵¹

'The apparent space between the gross projectivism of Galileo and a more sophisticated projectivism which allows for the 'new creation' understood as a manifest realm of response-dependent features is merely apparent. The initially promising variant on Projectivism is actually an unstable compromise' (Johnston 1998, p. 12)

In the light of this claim, it is natural to wonder whether Johnston still endorses his earlier claims about the response-dependence of values, accessible truth and theory acceptance, which gave response-dependence important work to do.

3. Similarities and differences

In this section, I'll sum up the similarities and differences between Wright's and Johnston's accounts of response-dependence, and place some other authors on response-dependence with respect to the issues discussed, thus creating a sort of map of positions (though without going into detail about the views).

Response-dependence of concepts or subject matter?

Wright (like Pettit) formulates his view as a view about certain kinds of *concepts*, while Johnston mostly puts his view in terms of *properties* (though he talks about

⁵¹ Johnston (1998), p. 11.

concepts and ‘conceptual identities’ in his early papers on response-dependence). However, this difference is less crucial than it might have seemed. Essentially, Johnston’s and Wright’s accounts are both theses of response-dependence of subject matter, not only of the concepts involved. In this, they differ from Pettit’s version of response-dependence (and Haukioja’s (2000) development of it), but side with most other authors on response-dependence. A notable example is Wedgewood (1998), who argues that response-dependent properties are those with response-dependent essences. Another is DeClercq, who aims to capture the distinction between response-dependent and -independent properties in a way that, unlike Wright’s, makes morality come out as response-dependent as expected. On DeClercq’s view, ‘a property is response-dependent [...] if there is an a priori guarantee that *if* the property is instantiated, we will come to possess a concept of it in due course’⁵² (i.e. at some point during the history of humankind). The idea is that such a close connection between concept and property will obtain only for response-dependent subject matters.

Lopez de Sa’s (2003) view is a mixture between the two paradigms; he thinks that response-dependent concepts with *flexible* specifications of subjects and C-conditions go with response-dependence of subject matter, while those with *rigidified* specifications go with low level location (though he doesn’t put it in those terms). On his view, the latter but not the former sort of account vindicates realism.

Response-dependence and realism

Johnston and Wright share the motivation outlined in Ch. 1: reconciling subject-relatedness with a moderate realist stance towards the subject matters under discussion. More precisely, for response-dependent domains, they give affirmative answers to the realism questions about 1) truth-aptness, 2) epistemic access, and 3a) existence of the disputed subject matter. Question 3b) on subject-independence is answered in the negative, but in slightly different ways. Johnston’s view is, of course, that response-dependence means that the properties in question are response-dispositions. Wright’s view is that the properties in question are merely deflationary, and have their extensions determined by the extensions of criterially

⁵² DeClercq (2002), p. 162, his italics.

governed concepts with responses in C-conditions as main criterion; best opinions determine the extensions of both concepts and properties. On Johnston's view, by contrast, there is a slightly more substantial referent in the picture, even if this referent is not a low level property, but a response-disposition.

With respect to the question of realism, Wright's and Johnston's versions differ strongly from Pettit's, which is consistent with low level reference, and thus with realism in more respects, or degrees, than Wright's and Johnston's versions (more on this shortly). Wright's and Johnston's views both amount to high level locations. While low-level properties may have a role to play, e.g. a causal one, they are not essential for the concepts, and are not their referents.

(As already stated, Lopez de Sa distances himself from Wright and Johnston on the issue of realism and subject-relatedness: for response-dependent concepts with rigidifiers in the specifications of subjects and C-conditions, we get realism, but not (any interesting) subject-relatedness, and no response-dependence of subject matter, but rather camouflaged low-level reference. For those with flexible specifications, we get response-dependence of subject matter (and subject-relatedness), but no realism. Powell (1998) also argues that realism is incompatible with response-dependence of subject matter, though see Wright (1998) for a convincing response to his criticism.)

Equations and conditions

Wright and Johnston employ similar materials to capture their distinctions, but also differ on some points. One point of difference is the shape of the equations. Johnston, like most other authors on response-dependence, employs basic equations or something close, whereas Wright relies on provisional equations in the hope that these will be immune to conditional fallacy problems. Johnston hopes to solve these problems by formulating the equations in overtly dispositional terms, while Wright sticks with subjunctive conditionals. (Other authors are more or less equally divided between the two options.)

As to the conditions that the biconditionals must meet in order to be a sign of response-dependence, Wright and Johnston agree on the almost⁵³ universally

⁵³ Except for Miscevic (1998).

accepted a priority and substantiality conditions. Johnston supplements this with a necessity condition, whereas Wright adds the independence and extremal conditions in order to capture the distinction he wants. These differences are not too important, though; the differences in philosophical content, not those in formal apparatus, are what makes the real difference between the accounts.

Applications

Lastly, there are differences with respect to intended application areas. Wright and Johnston agree (with most others) in using colours as a paradigm case of response-dependence. Johnston thinks of values as response-dependent, while this domain doesn't fit the requirements of Wright's version (though he admits that a response-dependence view of it would be attractive). Johnston's application to accessible truth does not have a parallel in Wright, except in the uninteresting sense that Wright, too, takes some truths to be dependent on responses. Wright's applications to intentional states and rule-following, on the other hand, do not have a parallel in Johnston. On the whole, the missing explanation argument probably makes Johnston more sceptical towards the idea of widespread response-dependence than Wright (pace Johnston's response-dependence thesis of accessible truth), though both agree that global response-dependence is an unpalatable and strongly anti-realist conclusion.

The differences between Johnston's and Wright's accounts are all relatively minor compared to the differences between these views and Pettit's, which is the subject of the next chapter. With the results from that chapter in place, we shall return to response-dependence of subject matter in Ch. 6.

Ch. 5: Pettit's ethocentric story and its potential

This chapter explores Philip Pettit's version of response-dependence, which is very different from Johnston's and Wright's – so different that their sharing the term 'response-dependence' may be more confusing than useful. In Pettit's version, response-dependence is a property of concepts only, and is neutral on the nature of the corresponding properties, objects, or states of affairs. The core of the proposal is an account of concept acquisition which was proposed in response to the Wittgensteinian problem of rule-following. In the first half of the chapter, I present Pettit's views. In the second half, I apply Pettit's account of the concept-acquisition of individuals to the concept-acquisition of entire communities. This yields an account of concept evolution that promises answers to some hard questions about concepts, linguistic practices, reality, and the relation between them. I discuss some of the prospects and problems of the account, and its relationship to Pettit's account.

1. Pettit's ethocentric story of concept possession

Philip Pettit's account of response-dependence was developed as a response to the Wittgensteinian problem of rule-following. The first version of the proposal, in Pettit (1990), focuses on the question what is required for something to be a rule, and goes on to argue that all appropriate requirements can be met by a response-dependence account of concept acquisition. But I will leave out the detail of this first proposal, since a thorough discussion of rule-following is beyond the scope of this chapter, and the aspects of Pettit's proposal most relevant for my purposes are different ones. In Pettit's later writings (e.g. 1991, 1998a, 1998b and 2005), a slightly different version of the account is given, this time addressing a different, though related, question about rule-following: *How can a finite set of examples, accessible to a finite mind, enable a subject to grasp a rule with a potentially infinite range of applications?*

Pettit's response to this problem gives centre stage to certain response-dispositions of subjects. The idea is that even if a finite set of examples is in principle compatible with an infinity of different rules – most of them strange and grue-like –

the examples may well exemplify a particular rule *to a subject*. (Or at least they might suggest a particular way of proceeding in the range of cases the subject would ever come across.) What makes this possible is the subject's response-dispositions and the role they play in the way we come to possess concepts.

About this, Pettit tells the following story: Some concepts are acquired by way of other concepts, by explicit definitions, explanations etc. But 'on pain of circularity', as Pettit puts it,¹ this can't be the case for all concepts. There has to be a class of basic concepts that are acquired in another way, by ostension or something similar. Pettit stresses that the basic concepts need not be the same for everybody, or even for a single individual over time. For example, people with normal vision may acquire colour concepts by ostension, whereas someone who is blind from birth will know them only by way of other concepts.²

How are basic concepts acquired? The first core element of Pettit's account are certain simple extrapolative dispositions: dispositions to notice certain salient similarities between samples, and to extrapolate from the samples in a certain way on the basis of this similarity. Suppose for the sake of the argument that 'red' is a basic concept. Then the idea is that exemplars like ripe tomatoes, blood, and British post-boxes simply and primitively seem to beings like us to have something in common. When they are pointed out to a learner and all given the label 'red', she will come to associate the concept with the common factor that seems salient to her, whatever its exact nature might be. Her reaction to new cases will be to apply the term to things that are similar to the original samples in the same way. Our learner need have no idea what the similarity consists in, or even conceptualise the fact that there is a similarity between the samples. Also, she need not be conscious of the particular way the samples affect her, e.g. that they give her a certain visual experience, nor must she have a term for that experience. All that matters is that she will be disposed to apply the term to new cases to the extent that they strike her as similar to the original examples in the right way.

¹ Pettit (1995), p. 11.

² Pettit (1995), p. 11. It may perhaps be questioned whether a blind person has the same colour concepts as normal-sighted people.

(In what way? Not by sharing a property or falling under a concept that is already familiar to the subject, say from her 'language of thought'; this would amount to re-introducing the Augustinian picture of language that Wittgenstein targets in *Philosophical Investigations*.³ The story must be that if a similarity is perceived that generates the required constancy (see below), then a concept can form, *regardless of the source of the perceived similarity*. The similarity must be taken as primitive, and there must be no hypothesis that it is due to recognition of a pattern already conceptualised within a language of thought.)

As is familiar from Kripke's Wittgenstein,⁴ a 'series' defined by a limited set of samples can in principle be continued in many ways, most of them strange and grue-like. The 'correct' way to extrapolate from a finite set of examples is underdetermined by the samples themselves. Pettit's point, however, is that the response-dispositions of a subject may pick out one of them as the natural way to go in all cases that the subject might come across.⁵ It is uncontroversial that such extrapolative dispositions exist; confront a child with a set of red objects, and she will get the idea and learn the concept in the way intended, and, once having got the idea, she will normally be in no doubt how to extend the concept to new cases.

Another familiar point from Kripke's Wittgenstein is that dispositions in themselves won't do the trick.⁶ Rules cannot be constituted by whatever we're disposed to do in each instance. First, 'rules' thus constituted couldn't be normative. Secondly, people are disposed to make (what is best described as) mistakes, e.g. forgetting to 'carry' in addition tasks. A distinction is needed between what seems right to a subject and what is right. As on Wright's account, this distinction is provided by the C-conditions. Pettit's account of C-conditions is different from

³ Wittgenstein (1953), especially §1-6 and 32.

⁴ Kripke (1982), especially p. 8-22.

⁵ There may still be a problem with cases that no-one would ever come across, e.g. additions of extremely large numbers. Pettit seems content to conclude that as long as the extrapolative dispositions cover all the cases that a subject might come across, this is sufficient for the solution to work. But for domains like mathematics, this might seem unsatisfactory.

⁶ Kripke (1982), p. 22-37.

Wright's (and Johnston's), though. We shall look at it in more detail in Ch. 7, but here is a quick summary:

In some cases, the verdicts based on 'salient similarity responses' are different for different people, or differ over time for the same person. In case of such 'inter- and intrapersonal discrepancies', a second disposition comes into play: the disposition to try to minimise such discrepancies by discounting circumstances that tend to yield conflicting judgements, and to think of judgements made in such circumstances as 'not counting'. In the face of unexplained disagreement, we are disposed to suspend judgement and seek explanations of deviant verdicts in terms of perturbing factors that somehow mess up the normal course of things in the problematic cases. By participating in this practice and having her verdicts corrected by (those of) others, and by those of her later selves, the learner acquires a sense of conditions that tend to yield deviant results, and comes to treat verdicts reached in such circumstances as unreliable and not to be counted. Likewise, she may identify circumstances particularly favourable to making judgements on the matter in question. The C-conditions emerge from this search for constancy as the ones that survive the discounting practice (or, in some cases, conditions that turn out to be particularly conducive; these might serve as basis for idealisations). The C-conditions are given by the way the practice works. As with the 'salient similarity' described above, it is not required that the subjects are consciously aware of the C-conditions or the discounting practice that gives rise to them; they need only proceed in the appropriate way, which theorists can describe in the way just sketched. Accordingly, Pettit describes this second order disposition as being as primitive and blind as the first; this is just the way we do it. However, there is a rationale for this way of proceeding: If you assume that the salient similarity is there because the cases genuinely have some property in common, and that other people track the same property as you do, then it is sensible to think that something has gone wrong if their judgement and yours get out of step with each other. In such a situation, the reasonable thing to do is to suspend judgement and try to find out what has gone wrong with the verdict of one or both parties in the disagreement.

These, then, are the two core elements in Pettit's account: the disposition to notice a salient similarity between sample instances, and the disposition to seek inter-

and intrapersonal constancy in the use of the concept, and to discount conditions that yield deviant results. Pettit calls the account ‘the ethocentric story’, from the Greek word Ethos (custom/ process of correction), which is thought to capture both core elements in the story: habit of response and practice of self-correction.⁷ To be response-dependent in Pettit’s sense, a concept must be acquired in the way the ethocentric story describes, or be defined in terms of concepts so acquired. (This means that response-dependence is a global phenomenon, which presumably it must be in order to address the problem of rule-following; we shall come back to this consequence shortly.)

1.1 Response-dependent concepts, response-independent properties

Concepts acquired in the way described by the ethocentric story (from here, ‘ethocentric concepts’) may well pick out things which have more in common than their disposition to elicit the response that makes us classify them together.⁸ Just as the characteristic water criteria point us to the natural kind H₂O, the salient similarities we notice between things that fall under a response-dependent concept can point us to common factors on lower levels, and in such cases it might be appropriate to view those lower-level properties as the referents of the concepts. For example, similarities in heat sensations turn out to point us to similarities in mechanical energy on the molecular level. We might choose to conclude that this property is the proper referent of ‘heat’. (Whether we should do so is another matter; there is also something to be said for the primacy of a phenomenal concept of heat, or for a view that two concepts of heat are at play simultaneously. But let us assume for a moment that the lower level location is the appropriate one.) Concepts formed in a response-dependent way might thus refer to response-independent properties and kinds. At first, a learner of a concept might have no idea what the corresponding low-level property is. For example, she might come to possess the concept of heat while knowing nothing about molecules (like children usually do, and like everybody did

⁷ Pettit (1991), p. 601.

⁸ Most of the points made in this section actually make more sense in the context of the community-wide version of the ethocentric story to be developed in the second part of this chapter. But as they also form part of Pettit’s view, they will be presented here.

some centuries ago). It might be that all we can say is that the exemplars have some property in common, and that this property is what we refer to by the word 'heat'. In this respect, the situation is exactly analogous to the case of 'water'; we might possess the concept of water while not knowing about H₂O. So response-dependent concepts can be referentially governed concepts, and the detail of the extension-determination can be deferred to substantial properties, if suitable candidates exist.

In other cases, it might turn out that there is no interesting underlying similarity at all across a group of instances that are classified together by way of a salient similarity – i.e. that all that unites them is their eliciting responses of a certain kind, and thus being classified together. In such cases, it may be appropriate to think about the concept as a higher level one – e.g. that it refers to a response-disposition. The account, in short, is neutral on the location issue.

Response-dependence in Pettit's sense is thus a property of concepts, but not of the properties, objects or states of affairs the concepts refer to. That a concept is acquired by way of response-dispositions in subjects does not entail that responses or response-dispositions make an ontological difference to the referents of the concepts; these can be as objective as anyone could wish for. In this respect, Pettit's account differs from accounts like Johnston's and Wright's that entail a high level location of the disputed phenomenon. Pettit-style response-dependence is compatible with response-dependence in Johnston's or Wright's sense; nothing in the account rules out that response-dispositional or Euthyphronic concepts might be acquired in the ethocentric way. In fact, the ethocentric story looks like a very plausible account of how such concepts may be acquired, and provides useful tools for understanding response-dependence of subject matter (see Ch. 6). But Pettit takes low level reference to substantial properties to be by far the most interesting case, and seems generally sceptical about (what in our terms are) high-level locations. Even for colours – everyone's paradigm example of response-dependence – he favours a low level location.⁹

⁹ Pettit (1998b), p. 65.

1.2 Biconditionals and conditions

Given these crucial differences between Pettit's version of response-dependence and the versions that take a stance on the location issue, it may come as a surprise that Pettit sticks with the familiar formal apparatus of response-dependence. He endorses the 'standard package' of an a priori and substantially specified biconditional. The exact shape of the biconditionals varies in his papers; in Pettit (2005), it is this:

- (1) For a given concept T, it is a priori that something is T if and only if it is such as to seem T in favourable conditions.¹⁰

Pettit's (use of the) biconditionals differ from Wright's and Johnston's in a number of respects. First, the *responses*. The phrase '*seems T*' allows for a variety of different responses that might help make a similarity pattern salient to us – visual experiences, taste, smell, sound, all kinds of bodily sensations, emotional responses, and presumably even the bodily responses ruled out by other authors. The 'seems T' formulation also allows for combinations of responses, e.g. combinations of looks and tactile experiences, as the relevant response for a given response-dependent concept.¹¹ According to Pettit, the one requirement on the responses is that they have to be relatively primitive and 'capable of doing the job required of them without being conceptualised'.¹² Thus, interestingly, the only response that doesn't seem to fit the bill is judgements – the very response that is considered paradigmatic on Johnston's account, and capable of doing the job alone according to Wright's.

Next, the *favourable conditions*. As already explained, the ethocentric story provides an account of how they are chosen: The C-conditions are those that survive the practice of discounting conditions that yield inter- and intrapersonal discrepancies in the use of the concept. Pettit adds, in agreement with Wright and Johnston, that the favourable conditions should be non-trivially specified. In 'A theory of Normal and Ideal Conditions' (1999), he develops a functionalist account

¹⁰ Pettit (2005), p. 11.

¹¹ See Pettit and Jackson (1998), p. 109-11, for arguments that some response-dependent concepts depend on combinations of different responses.

¹² Pettit (1998a), p. 120.

of C-conditions designed to ensure that the a priori and substantiality requirements can both be fulfilled. I shall treat his proposal in more detail in Ch. 7.

A significant difference between Pettit's and other accounts of response-dependence is *the role of the biconditionals*. Everyone agrees that the biconditionals can't be *analyses* of the disputed concepts; the 'echo' in the biconditionals prevent that. The role of the biconditionals consequently can't be to *dictate* the application conditions of the concepts. Pettit explains further that on accounts like Johnston, the biconditionals do something close: they *reflect* the application conditions for the concept. On Pettit's account, the biconditionals have nothing to do with application conditions. Rather, they reflect the *possession conditions* of the concepts in question; it reflects the way concepts users acquire expertise with the concepts based on a particular sort of responses and the discounting practice that gives rise to the C-conditions.¹³

A final, important difference between Pettit's and other accounts is *the source of the a priori of the biconditionals*. On Wright's, Johnston's, and similar accounts, the biconditionals are a priori in virtue of the close connection between the disputed properties and responses in C-conditions. On Pettit's view, however, response-dependent concepts can refer to response-independent, substantial properties that are only contingently connected with responses in humans. How can it be a priori that such a property will be correlated with responses in C-conditions? Pettit's answer is that the biconditional is a priori as a result of the way the reference of the concept is fixed. For a property to deserve the job as referent for the term T, it must be the (possibly disjunctive) property that gives rise to the right kind of responses in C-conditions. This makes it a priori that the property T refers to – whatever its nature might be – will give rise to such responses. Analogously, even if 'water' refers to the natural kind H₂O, which is only contingently connected with watery appearance, it can be a priori that

¹³ Pettit (1991), p. 602. The terminology of possession conditions is not prominent in his recent writings, but still fits well with the views stated there.

- (3) something is water iff it is the stuff that is actually dominantly causally responsible for watery appearances (please substitute your favourite reference-fixing description for ‘water’),

given that this is what fixes the reference of the term ‘water’, and hence a property has to fit the description in order to be the referent of the term.

The point about the reference-fixing role of the a priori biconditionals raises some problems that will be discussed in Ch. 6. One problem concerns the possibility of reference failure. If a term purports to refer to a substantial property (a natural kind or similar), there is always a risk that it misses its target because there simply isn’t a suitable property that fits the reference-fixing description (near enough). Suppose we have a response-dependent concept that purports to refer to a substantial property. There is a stable practice with the concept, and a set of C-conditions in which we get the required amount of inter- and intrapersonal constancy. But on closer investigation it turns out that there is no suitable low-level property to serve as the referent of the term. In such a case, will the concept fail of reference? If it does, there is a worry that the biconditional may fail in the right to left direction. There may well be things that are such as to seem T in favourable conditions and consequently make the right hand side true, but if ‘T’ does not refer, the left hand side will presumably be false or meaningless. Furthermore, if the biconditional holds out such a hostage to fortune, how can it hold a priori even for concepts that in fact turn out to refer, given that we know about the referents only a posteriori?

Secondly, there is the problem that if the detail of the extension-determining is deferred to a substantial property, there is a risk that some few instances of that property would not seem T no matter how conducive the conditions. More on these two problems in Ch. 6.

1.3 Realism and response-dependence

Pettit (1991) argues that response-dependence in the sense of the ethocentric story is compatible with realism except for a few minor compromises. More precisely, it is compatible with the following three theses that, in Pettit’s view, capture the central tenets of realism:

- *the descriptivist thesis* that the concepts in question are used to make (truth-apt) claims about a distinctive range of entities, properties or states of affairs (realism question 1 from the taxonomy in Ch. 1)
- *the objectivist thesis* that those entities/ properties/ states of affairs exist, and do so independently of our speaking or thinking about them (our 3a and 3b), and
- *the cosmocentric thesis* that ‘the area tracked by the vocabulary is territory where human beings enjoy no guarantee of epistemic success, being capable of pervasive ignorance and error’¹⁴ (something close to our question 2 about epistemic access).

Pettit-style, ‘ethocentric’ response-dependence is compatible with the descriptivist thesis, as it doesn’t entail that the statements involving response-dependent concepts are not truth-apt, or that the properties, objects or states of affairs they posit must be reducible to those of another domain. It is also compatible with the objectivist thesis, as it does not entail that the properties etc. in question do not exist, or even that they do not exist independently of us. Even if we latch on to a property via certain responses, the property itself can be as real and as independent of us as one could wish for. Finally, Pettit-style response-dependence is compatible with the cosmocentric thesis in the most important respects; it does not entail that the subject matter is anthropocentric in the sense of precluding ignorance and error. There is a small compromise in that error and ignorance in C-conditions is ruled out.¹⁵ But the compromise is very limited. First, it only applies to basic concepts. Secondly, for any given case it could always turn out that the C-conditions were not fulfilled after all, so we never get an inappropriate a priori guarantee to have got things right.

Pettit argues further that response-dependence is consistent with two core ideas behind the cosmocentric thesis. The first he calls epistemic servility – roughly, that we are in the business of trying to get things right, and that whether we do so is determined by factors outside our control. The second is ontic neutrality – that the

¹⁴ Pettit (2005), p. 12.

¹⁵ Though see Holton (1991) for arguments that not all response-dependent concepts rule out ignorance and error in this way.

properties or objects in question could be accessible to beings very different from us, and hence are not closely tied to a human perspective.¹⁶

Response-dependence does have a couple of surprises in store for the realist, though. One is a possibility of indeterminacy in the concepts if response-dispositions don't deliver a clear verdict in all cases (e.g. in borderline cases of vague concepts such as 'bald' and 'tall'). Another is that there will be more than merely inductive connections between the properties to which some response-dependent concepts refer and certain typical human responses. For example, direct recognition of the property of cruelty will arguably always be accompanied by a feeling of aversion, and such a close connection might not have been expected by a realist who thought that cruelty was a property that was independent of human responses.

In all, there is no big conflict between response-dependence and realism. In Pettit's words,

people's responses do not shape certain things so that they fall under the concept of redness, they shape the concept of redness so that it falls upon those things. (Pettit 1991, p. 622-23.)

A consequence of the compatibility with realism is, of course, that Pettit's response-dependence distinction doesn't tell us anything interesting about the realism-related differences between different sorts of concepts. As part of a project like Wright's, it would thus be of limited use.

The conclusion that response-dependence in Pettit's sense is not in serious conflict with realism is a welcome one, especially because Pettit's brand of response-dependence is intended to apply globally. In order to work as a solution to the rule-following problem, it must work across the board; all concepts must be response-dependent in the sense of either 1) being acquired in the way described by the ethocentric story – this goes for basic concepts – or 2) being a non-basic concept based on ethocentric concepts. The claim is not that all concepts are acquired in the way described by the ethocentric story. This would be implausible, as in many cases it would be hard to find a suitable 'salient similarity' response (for example, what is

¹⁶ Pettit's notion of ontic neutrality brings to mind Wright's Wide Cosmological Role constraint. Wide cosmological role is presumably a (fallible) symptom of ontic neutrality.

the typical response to something's being a mantelpiece?). Pettit's claim is only that all *basic* concepts are acquired in this way. (He remains neutral on which concepts are basic, apart from the claim that they may be different for different people.) However, if being a basic, ethocentrically acquired concept entailed some sort of anti-realism, and if that anti-realism was infectious to concepts acquired on the basis of basic concepts, on the model of the globalisation arguments mentioned in Ch. 4, this might have spelled trouble for the account.¹⁷

The compatibility of Pettit-response-dependence with realism makes Pettit's response-dependence-based attempted solution to the rule-following problem seem more promising than Wright's if both would do the job (I'll refrain from taking a stance on whether they would). The accounts are built from very similar materials, and share many features: C-conditions that can be assumed fulfilled until evidence appears to the contrary, the appeal to C-conditions in order to get a distinction between seeming F and being F, and the distinction between basic and non-basic concepts (Pettit) and a somewhat similar distinction that Wright draws between basic and non-basic judgements.¹⁸ But Wright-style response-dependence has a strong anti-realist flavour, and this will be particularly problematic if a response-dependence thesis for meaning turns out to entail global response-dependence. If both work, Pettit's account is preferable because it avoids such a consequence.

¹⁷ It is not clear to what extent non-basic concepts acquired on the basis of ethocentric concepts would be infected with any anti-realism affecting the basic concepts. We cannot assume without argument that any limitation on the realism level for basic concepts carries over to the rest. For example, it seems possible that all the ethocentrically acquired concepts are response-dependent in something like Johnston's and Wright's sense, and that all referentially governed concepts – concepts that defer the extension-determining to independently existing properties – are non-basic and introduced via such concepts (tasteless, odourless liquid etc.). I.e. it could be that all ethocentric concepts are closely tied to a human perspective, but enable us to acquire concepts that are not. Thus, it is not clear what features will carry over from the basic to the derived concepts.

Relatedly, it is not clear what sense of response-dependence would carry over from basic to non-basic concepts. This makes it unclear what Pettit's claim of global response-dependence really amounts to.

¹⁸ Wright (1986); the common beginnings of, and similarities between, the two accounts are very visible in this paper.

This concludes the exegesis of Pettit's views. In the next section, I shall consider some questions it raises.

1.4 Some questions about Pettit's account

Pettit presents the ethocentric story as a story about the concept acquisition of individuals. This raises a couple of questions. First, *is concept acquisition the kind of thing we could tell a meaningful story about a priori, or is it an empirical question?* If it is the latter, then a charge of illicit armchair science may threaten Pettit's account.

There is clearly a sense in which it is an empirical question how concepts are acquired. However, there are important philosophical questions in the vicinity, such as these: 'Given the sceptical challenge from Kripke's Wittgenstein, concept acquisition looks impossible. Can we tell a story about how it could have come about?' and 'What is conceptual content, and what does it take for language users to access it?' Pettit's ethocentric story suggests answers to such questions – in my view, promising answers. If viewed as an in-principle-story about how concept acquisition is possible, the account makes a lot of sense, and does not attempt to do anything that should not be possible from the armchair.¹⁹ (A relevant analogy may be Hobbes's in-principle-story about the state of nature in *Leviathan*.²⁰)

A second question is this: *Is the story best understood, and the interesting questions best addressed, by a story about the concept acquisition of individuals?* This time, I think the answer is negative. The most interesting version of the ethocentric story is a story about the concept acquisition of *entire communities* – in other words, a story about concept evolution. I will sketch such a story below and discuss the advantages it would have, especially when seen through the lens of the

¹⁹ On the legitimacy or otherwise of such in-principle-stories, see a footnote later in the chapter.

²⁰ Hobbes (1985). Hobbes takes as starting point an imagined 'state of nature', in which everyone is at war against everyone else, and argues that people are better off relinquishing some of their rights and power to a sovereign which can make co-operation and peace possible by securing the keeping of covenants between people. The 'state of nature' is not supposed to be historical reality, but is a thought experiment designed to prove Hobbes's points.

levels framework. It is not clear whether Pettit himself would accept such a story, or how it relates to his account. We shall discuss the relationship between his account and the ethocentric story of concept evolution at the end of the chapter.

2. The ethocentric story applied to concept evolution

The ethocentric story of concept acquisition for entire communities would be analogous to the version for individual learners: Some concepts come into being by introduction and definition in terms of other concepts. But this can't be the case for all concepts, or linguistic rule-following would never have got off the ground in the first place. Some concepts – let's call them basic concepts – will have to have come into being in a different way. As in the 'individual version' of the ethocentric story, the story would be that basic concepts owe their existence to the response-dispositions of subjects: dispositions to notice salient similarities, and to seek constancy in verdicts based on them. Basic concepts come into being because people's primitive (non-conceptual) response-dispositions make things like ripe tomatoes, blood, and heated metals appear saliently similar. If the perceived similarity patterns are stable enough across different people, signs (such as the term 'red') can be introduced to communicate about things that are found to appear in the characteristic way. In virtue of people's similar dispositions to see salient similarities, such signs will work as means of communication, even while it is not yet known what similar-seeming cases have in common apart from appearing similar in a particular respect.

In certain cases, discrepancies arise between the verdicts of different people, or between the verdicts of the same person at different times. Things that would normally appear in the characteristic way – e.g. look red – appear differently in the dark, or in differently coloured surroundings, or when in rapid motion. In such cases, the second order disposition to seek to minimise discrepancies kicks in: the group endeavours to avoid discrepancies by discounting circumstances that yield deviant verdicts, and by explaining away deviant judgements as a result of perturbing factors ('colour appearances in the dark do not count' etc.). In this way, the C-conditions come into play, and with them the distinction between seeming red and being red that

is required for the emerging linguistic rule to function as a normative constraint on the practice with the concept.

The relevant response-dispositions – both the first order disposition that makes cases appear similar and the second order disposition to seek to avoid discrepancies – have to be shared among the relevant group in order for the account, and the practice it describes, to work. But given that the members of the group belong to the same species, there is a good chance that this will be the case. As Daniel Nolan puts it (in a different context),

Across times and across cultures, human beings make similar similarity judgements in many different ways; ask anyone to pick the odd one out from two badgers and the Taj Mahal, and they'll be able to do it! It is probably vital to our being able to learn anything that we make fairly stable similarity judgements, and learning a language would seem to require that we spontaneously make similarity judgements that are approximate to those of others in our community.' (Nolan 2005, p. 23.)

This story suggests promising answers to some hard questions about concepts, linguistic practices, and the properties, objects or states of affairs that language is used to talk about. We shall focus on the following questions:

- How can concepts come into existence, be used, and be in good standing before the exact nature of the referent, or even its level, is known?
- How can concepts made of 'for us'-material come to refer to independently existing properties and kinds?
- Can we tell a unified story about how different kinds of concepts – natural kind concepts, concepts of response-dependent subject matters, and concepts (if any) for which relativism is appropriate – come into being?

These interrelated questions set the stage for the discussion below; I shall go through all the components needed for answering them, and then state the suggested answers explicitly in the summary at the end of the chapter.

2.1 Some constraints on reference

In order to fully answer the advertised ‘hard questions’, we need a few preliminaries. First, we need to consider the following question: *What does it take for a property (or object, etc.) to be the referent of a concept?*

This is a difficult question, and here I can only offer some very preliminary considerations which are especially relevant to our discussion. A part of the answer is presumably *correlation*; in order for a property (or object, etc.) to be the referent of a concept, there must be a reasonable degree of correlation between instantiation of the property and application of the concept.²¹ The referent of an ethocentric concept should be sought among the properties that are correlated with its use (or near enough; for referential concepts, a certain amount of divergence between actual use and the extension of the properties must, of course, be allowed). Another part of the answer may be that certain *counterfactual connections* are required: that, other things equal, if the property had not been present in a given case, the concept would not have applied, and had the property been present in a case where it’s not, the concept would have applied.²²

A third part of the answer might be that there must be a *common factor* on the level on which the reference of the concept is located. This is controversial, however. To me, it seems natural to locate reference on levels where a common factor is found – even if the only common factor turns out to be satisfaction of a set of criteria. But many disagree, and locate the reference of concepts on lower levels than where the common factors are found. For example, (some) colour physicalists locate colours with surface structures, even if the common factors are found on the level of dispositions to elicit colour experiences, and to a lesser degree the dispositions to reflect light of certain wavelengths. What motivate such locations are usually ideas

²¹ At least this would go for ethocentrically acquired concepts; there may be concepts introduced by some sort of stipulation/ definition, for which users are completely mistaken about what things, situations, etc. the concept applies to.

²² If a concept is not thought of as a rigid designator, but as referring to different properties in different worlds, then these counterfactual connections will be limited to worlds in which the referent is the same. An assumption that reference is rigid underlies much of the reasoning in this work, though I will not argue for it here.

about the low level as ontologically privileged over the higher level – as the levels on which the real properties are found. This picture relies heavily on the idea of a basic level discussed in Ch. 3. If this presupposition turns out to be unwarranted, the motivation for location on levels lower than where common factors are found presumably goes missing. In the following, I shall assume without further argument that common factors are an important guide to reference.

What other factors influence the question on which level the right location for a concept is found? We have already considered some candidates in Ch. 2: Beside empirical information about correlated factors and common factors, *intuitions* about the concepts and their subject matter arguably provide clues to location. Such intuitions could be explicitly about location – e.g. intuitions that the concept refers to a natural kind or similar. Or they could be intuitions that bear on the location question in more indirect ways, e.g. intuitions about causal efficacy (presumably pointing to low level locations) and transparency intuitions (pointing to high-level locations). These intuitions could be acknowledged by competent users of the concept, or they could be implicit in the practice for theorists to discover, but not acknowledged by participants.

(Finally, of course, for concepts that are not ethocentrically acquired, but introduced by *definitions* or *stipulations*, these can impose demands on what a suitable referent must be like.)

2.2 The open location question

Pettit's ethocentric story explains how it is possible for a learner, say a young child, to acquire a concept without knowing the exact nature of its referent. The same sort of story can now be told of entire communities: a concept may come into use and be in good standing even if no member of the community knows the exact nature, or even the level, of its referent.

On the ethocentric story of concept evolution, only two things are needed for a basic concept to form: A similarity made salient by response patterns, and sufficient inter- and intrapersonal constancy in judgements based on the responses for a stable set of C-conditions to form and to provide a seems/is distinction. Accordingly, all a subject needs in order to acquire a concept is the disposition that makes the similarity

pattern salient to the subject, and the second order disposition that allows her to participate in the discounting practice in the right way.

None of these features prejudge the question of underlying features. The constancy in responses in C-conditions could be due to a lower-level common factor, as in the case of natural kind concepts (e.g. water, assuming for a moment that this is a basic concept). Or it could be due to a smallish disjunction of such features (as with jade). Or it could be due to a more complex set of features, as in the case of colours (in this case, the physical common factor is something as specific as a disposition to reflect light with certain distributions of wavelengths on three different bands of wavelengths that the three types of receptors in the human eye are receptive to). Or, finally, the constancy could be due to pure coincidence, with no interesting story, however complex, to be told about low level features.²³ (But still, the story allows underlying features to influence the extensions of the concepts; more on this in the next section.)

This means that it can be left open whether the high level similarity is correlated with a low level similarity, and whether such a low-level similarity would deserve the status as the referent of the concept. This allows a concept to be established in the ethocentric way before it is known on which level an appropriate referent is to be found, or what its exact nature is. These questions can be left for later empirical investigation, and the practice around the concept can do just fine without an answer.

This story is very similar to the one told in section 1.1 about how response-dependent concepts can refer to response-independent properties. The reasoning in

²³ Cases of the latter type are presumably rare, however, as is suggested by Pettit's example of 'ping' and 'pong' (1991, p. 615). These 'concepts' mean absolutely nothing. Yet there is a lot of regularity in the classifications people make when asked which things are ping and which ones are pong. For example, ice cream is almost invariably placed in the ping category, whereas soup is classified as pong. The explanation of these regularities is presumably similarities in the sound of the words or something of that sort, but it makes for a regularity explained in terms of underlying features all the same (though the relevant underlying features are a long way from being the kinds of properties that are appropriate as referents; this sets 'ping' and 'pong' apart from the colour case, where locating the reference of the concepts with the complex low-level candidate would not be a completely crazy move, and the case of water, where location with the low level common factor is a natural choice).

that section actually makes a lot more sense if based on the community version of the ethocentric story rather than the individual version. Strictly speaking, the conclusion in case of individuals will only be the fairly uninteresting one that a concept can be learned through ostension with the help of response-dispositions even though it is the concept of e.g. a natural kind, or perhaps a concept whose meaning is given by a definition in terms of other concepts. This much is fairly uncontroversial; whatever pedagogical methods work in language learning and teaching should be ok. When teaching a child the concept of *water*, pointing out samples in everyday contexts works a lot better than explanations about oxygen and hydrogen molecules. Likewise, when teaching mathematics, it may be preferable to let students acquire the concept *function* by examples and by doing exercises, rather than presenting them with inaccessible definitions. And when teaching the concept *cat*, exemplars will do just fine for ordinary purposes, and the story about DNA or whatever is essential to being a cat will be relevant mainly to biologists and philosophers. Relatedly, individuals can presumably master e.g. scientific concepts parasitically, as long as there are people in their linguistic community that know the full story about the concepts and the properties they refer to. This opens the possibility that the ‘ordinary people’ may acquire their (limited) mastery of the concepts in something like the ethocentric way, even though the experts acquire their more advanced mastery in a different way. But this possibility in itself says little about the relation between the concept and its referent.

The more interesting claim would be the community version of the same thought: that even entire communities may acquire concepts in the ethocentric way, even if (some of) those concepts subsequently turn out to refer to properties best described in another way than by the similarity pattern (e.g. by its chemical composition (‘water’), DNA (‘cat’), or definitions in terms of other concepts (‘function’)). This possibility is explained by the ethocentric story of concept evolution.

2.3 Accessing independently existing properties by concepts made of ‘for us’ material

Despite its neutrality regarding underlying features, the ethocentric story allows underlying features to have their say in determining the extensions of ethocentrically acquired concepts (often behind the backs of the participants in the linguistic practice). Low level features can influence the extension of an ethocentric concept, first, by being the normal cause of the relevant responses in C-conditions. In such cases, other things equal, the response-pattern, and hence the extension of the concept, would be different if the extension of the low level property (or complex of low level properties) were different. Secondly, and importantly, low-level features can influence where the C-conditions end up. If the regularity in responses is due to these responses tracking a low level property (or complex of such properties), then the circumstances that give rise to inter- and intrapersonal discrepancies in judgements will arguably be the circumstances in which the correlation between the salient similarity response and the underlying factor tends to break down. Discounting verdicts made in these conditions will facilitate tracking of the low level property by the responses, even if the nature (or even existence) of the low level property is not yet known. This would make it reasonable to conclude that a concept evolved in the ethocentric way could refer to a low level property – a property that has an extension determined independently of the concept or the practice around it.

In cases where the causal story is more complex, and no appropriate low-level reference candidate is available, there is still space for underlying features to do causal work that ensures the correlation between characteristic properties on different levels. When discrepancies arise in the judgements based on the characteristic response, there will usually be an explanation in terms of low level factors. For colours, for example, these (simplified) explanations might be given: ‘The object looks more yellowish than usual because it is placed in yellow light, and though its surface structure and reflectance spectrum remain the same, it reflects more yellow light than usual simply because the yellow light is what is there for it to reflect’; ‘This man has only two of the usual three receptor types in his eyes, so he lacks the red/green contrast, and hence his colour judgements in certain parts of the spectrum are not to be trusted’ etc.

Thus, even if a concept can get off the ground without presuppositions about the level of the referent, the story leaves lots of work to do for lower level properties. This would explain how we get the situation with relevant properties on several levels coinciding in C-conditions, but able to come apart in less favourable conditions and in counterfactual scenarios. In cases where a (well-behaved, i.e. not too disjunctive) low level property is the normal cause of the response, it may be appropriate to view that property as the referent of the concept, and to defer the details of the extension-determining job to that property.

This combination – the possibility of the concepts coming into being in a way that does not require users to be familiar with the underlying features, and the possibility that the high-level similarities lead us to, and allow us to track, low level features, makes it possible to address a challenge which might otherwise have seemed insurmountable: *How can concepts built of 'for us'-material come to refer to properties/entities/states of affairs that exist independently of our cognitive makeup and classificatory practices and dispositions?*

We might seem to face the following dilemma: If concepts are meant to refer to (low-level) properties that are constituted independently of our cognitive and perceptual capacities, classificatory practices and dispositions, we need an account of what could make our concepts refer to such properties, and how we could access the extensions of the concepts if they did. But how can we be sure that our connection with these properties is such as to make it appropriate to treat them as referents of the concepts, and how can we be sure that when the conditions for application of the concepts are met, the property is really present?

If, on the other hand, concepts are made of 'for us'-materials, so to speak, i.e. are tailored for our cognitive and perceptual makeup, response-dispositions etc., won't they mirror only classifications that are relative to a subjective standpoint? If our concepts depend too strongly on our classifications and responses or are purely consensual, they may not be very well suited for the purpose of learning about the world; there is a risk that knowledge in terms of them will be knowledge of our projections alone.

The ethocentric story offers a way out of this dilemma by showing how high-level similarities provide a means of access to low-level similarities (and with no

need to know the link between the two in advance). It thus offers an attractive account of how we can make epistemic and ontological ends meet. This is an important potential advantage of the account.

2.4 Location, intuitions, and low-level common factors

Many concepts come with presuppositions, direct or indirect, implicit or explicit, about the level of the referent. Some concepts come with intuitions that make a high level location the natural choice, e.g. intuitions about transparency and the essential role of appearances. Pain would be the obvious example; what makes something a pain is that it hurts, no matter what the low level story is. We shall have more to say about such concepts – the strongest candidates for response-dependence of subject matter – in Ch. 6.

Other concepts come with intuitions pointing to several locations at once, making the location question hard to settle (especially failing a clear empirical verdict, e.g. the discovery of the existence of, or total lack of, an appropriate low level reference candidate). As suggested, colour concepts presumably belong in this category, which is what makes the location dispute in that area so intractable. Yet other concepts may come with no location-relevant intuitions at all.

But for many concepts, it is natural to expect low-level reference – or, more accurately, the expectation is the more innocent one of unproblematic reference to a property in the world: a property that could be described as subject-independent, substantial, and probably low-level if we thought about the matter in those technical terms. Such concepts would presumably include *water*, *cat*, and presumably also *red* and other colour concepts, despite the competing intuitions pointing to a high level location. In fact, this may well be the default position for most concepts, as Pettit suggests.

If such a concept turns out to be correlated with a substantial, well-behaved low level property, this property can be thought of as the referent of the concept. But in other cases, such as colour concepts, no such property is to be found. Though we might have expected low-level reference – or at least unproblematic reference to objective properties – it turns out that the low-level correlates of colour concepts are very complex disjunctions of low level properties – arguably too disjunctive to serve

as referents.²⁴ If reference requires a common factor on the level on which the referent is located, colour concepts fail of low-level reference. On some views of concept formation and reference, this discovery would force a conclusion that the concepts were flawed, and should either be discarded altogether or replaced with new concepts suited for high-level reference. But the ethocentric story allows a less revisionary conclusion. When faced with the discovery of low-level reference failure for colour concepts, we may follow Johnston's lead and seek their referents among higher level properties, e.g. response-dispositions or the deflationary properties correlated with criterially governed concepts.

For many concepts, this picture seems plausible. Many concepts can be in good standing regardless what the situation turns out to be like on lower levels, even if they come with presuppositions in favour of low level reference if a suitable referent can be found. For many concepts, the natural response to low level reference failure is not an error theory, but a location on a higher level. The ethocentric story accommodates this possibility, neutral as it is on the location question. The core part of the account of the concept, the story about the salient similarity and the discounting practice, can remain in place, while all that needs to be adjusted is the presupposition of low level reference (plus, of course, any intuitions that might motivate it). This allows a less revisionary account than many alternatives (including Johnston's view that concepts are defined by the core intuitions about the subject matter in question).

Intuitions about location are a diffuse lot in any case. They tend to vary a lot across people and times, and are highly susceptible to philosophical manipulation. So it is an advantage of the account of concept evolution that it allows a concept to survive the discovery that some of the intuitions connected with it cannot be accommodated.²⁵

²⁴ Though see Jackson (1998), p. 105-112, for an argument to the contrary.

²⁵ Some externalists would judge intuitions about location altogether irrelevant to the question about the reference of the concepts. To them, my remarks above about intuitions and reference will be of limited interest. However, even for externalists, the ethocentric story should be attractive. The dispositions to notice salient similarities and discount circumstances that yield deviant verdicts – blind

This puts us in a position to give a more complete answer to the question how a concept can exist and be in good standing before the nature of its referent, or even its level, is known: The ethocentric story is neutral on the amount of regularity found on lower levels, and would work for both high and low level concepts. It also allows the location question to remain open for a given concept until any relevant facts about underlying properties are in. Thus, it explains how it is possible for us to refer to a property even if we don't yet know its exact nature, or even on which level it is to be found. This is a significant advantage of the story.

For similar reasons, an ethocentric story of concept evolution offers an account of how concepts could come to equivocate between levels in the way described in Ch. 2 – and indeed, how this could be the default position for ethocentric concepts. The ethocentric story allows underlying factors to influence the C-conditions in such a way that C-conditions will be the conditions in which characteristic features on different levels are correlated. Because of this correlation in C-conditions, there will often be no need to distinguish the features clearly, and nothing will force a 'decision' (implicit or explicit) about what level is most important for the concept. (Indeed, such a choice could be risky, since an unequivocal choice of a low level location would open a possibility of reference failure, and cut off the fallback option of higher level location.) So level equivocations are natural on the ethocentric story.²⁶ Many concepts arguably contain level equivocations; we have looked at a few (colour concepts and concepts of mental states) in Ch. 2. The ability to account for this phenomenon is another advantage of an ethocentric story of concept evolution.

2.5 The possibility of error theories

The account just given invites an objection regarding the possibility of error theories. In effect, the suggestion is that in the face of low-level reference failure, we can chase the referent up the levels until an appropriate candidate property is found,

as Pettit takes them to be – serve as a way of latching on to potential referents (underlying low-level common factors) that should be acceptable even to externalists.

²⁶ For the same reasons, an ethocentric story of concept evolution would make a perfect companion for an equivocation view as described in Ch. 2.

rather than conclude that the concept is in bad standing. If this picture is right, can a concept ever turn out to be in bad standing as a result of reference failure? If not, the account is in trouble; there are concepts for which an error theory seems the appropriate conclusion in the face of reference failure. An example is the concept of *phlogiston*. The term was introduced to name a natural kind characterised by negative mass, a certain role in combustion, etc. But it turned out that no such kind exists. In the face of this discovery, we did not seek a high level referent for the concept, but concluded that there is no such thing as phlogiston. How could that be if higher level location provides a fallback option?²⁷

The answer to this worry is that the fallback option is not available in all cases. It will be unavailable if the concept comes with too strong presuppositions about low level location, or if the concept would become useless without a low level referent, or if, as in the case of phlogiston, the concept is a theoretical postulate, for which it is stipulated that its referent, if any, is (what can be described as) a low-level property. In such cases, error theory is the appropriate conclusion in the face of low level reference failure. But in other cases, such as colour concepts, the appropriate conclusion is that the concept is a high-level one.

The contrast between the two kinds of cases can fruitfully be formulated in terms of a new bit of terminology (which will also be useful in later chapters). Let *level-rigid* concepts be concepts that are tied to a particular level and must refer to something on that level or not refer at all (for any of the reasons just mentioned). Phlogiston is a paradigm example. And let *level-flexible* concepts be concepts that permit location on different levels. Examples include colour concepts, and also the concept of heat; even if it had turned out that heat did not correspond to a well-behaved low-level property, the concept would have remained in good standing as a phenomenal concept.

The answer to the challenge about error theories, then, is that some concepts are level-rigid, and those concepts will be in bad standing if they fail of reference on the intended level. Level-flexible concepts, by contrast, can be in good standing despite low-level reference failure, as the referent may be sought on a higher level,

²⁷ The challenge regarding the possibility of error theory was given by Crispin Wright. The formulation in terms of the phlogiston example is my paraphrase.

and, if necessary, located with a deflationary property corresponding to a criterially governed concept.

2.6 Level-rigid and level-flexible concepts

The distinction between level-rigid and level-flexible concepts will be useful in the following chapters, so we shall put a little more flesh on the bones.

First, which concepts are level-flexible, and which are level-rigid? Level-rigid concepts paradigmatically include theoretical postulates like phlogiston and presumably most other scientific concepts. However, there are presumably also high-level-rigid concepts – concepts, e.g. pain, for which the essential thing seems to be ‘on the surface’, and to which the situation on lower levels is in a sense irrelevant (though maybe causally relevant). Euthyphronic concepts – response-dependent concepts in Wright’s sense – are concepts of this kind.

Ethocentric concepts are presumably either level-flexible or high-level rigid. According to the story told about them, such concepts can be in perfectly good standing without a low-level common factor; all that is required for them to work is the combination of salient similarity responses and sufficient constancy across people and times for a stable set of C-conditions to form and establish a seems/is distinction. So presumably no ethocentric concepts are low-level-rigid (though of course low-level rigid concepts could be defined on the basis of them). But they may well be high-level-rigid (e.g. Euthyphronic). If there are high-level-rigid concepts (as seems likely), an ethocentric story would be an attractive account of how they have come about. It might even be that all ethocentric concepts are high-level-rigid, though this view would take us very far from Pettit’s starting point. However, level-flexibility also seems an appealing option for ethocentric concepts. Until we have a grasp of which concepts, if any, are ethocentric, it is difficult to determine whether ethocentric concepts will be level-rigid or -flexible.

In general, the easiest way to determine whether a given concept is level-rigid or level-flexible is to ask questions like

- 1) ‘On what level does the referent belong?’ and
- 2) ‘What would happen if we discovered a lack of a referent on a given level?’

For a level-flexible concept, 1) should be an open question (unless the location question has already been settled, e.g. by the discovery of a low level referent), and the answer to 2) should be 'locate the reference somewhere else; the concept is still in good standing'. For a level-rigid concept, there should be a clear answer to 1), and the answer to 2) should be 'what a surprise; then the concept must be eliminated or changed'. (Of course, this test would only be useful in this very direct version if the level framework and the necessary terminology has been conveyed).

In theory, the distinction between level-rigid and level-flexible concepts is clear enough. But when applied to real-life concepts, it is probably better thought of as a continuum than a sharp, polar distinction. Many real-life concepts presumably come with expectations gravitating towards one level, and often with some uses that would be undermined by reference failure on that level, but still with a possibility of retreat to another if these expectations are disappointed. It may often be unsettled whether such a move is best described as a change in the concept or as a discovery that the referent of the concept was on a higher level than we expected. Also, it will probably often be unsettled how the concept is supposed to work, because the different levels involved are not distinguished, and the practice with the concept does not determine that one or the other location is right. This situation may be described as level-flexibility, but in some cases it may also be unsettled whether level-flexibility is the right conclusion, or whether the concept works in such a way as to force a high-level location. So it is reasonable to expect that many concepts will be hard to classify, and classifications will be controversial. This doesn't, however, make the distinction less interesting or useful.

What settles the location question for level-flexible concepts? It seems reasonable to take Pettit's line about ethocentric concepts and conclude that, other things equal, low level reference gets priority. If empirical investigation into lower level common factors has a role to play in settling the location question, as suggested by Johnston on colours, this would suggest that level-flexible concepts should be located on the lowest level on which an appropriate common factor can be found (at least in some or most cases).

2.7 Ethocentrism and relativist views

I have stressed as an advantage of the ethocentric story that it explains how a concept can exist and be in good standing while it is left open on which level the referent of the concept belongs. This section focuses on a closely related advantage: the ethocentric story offers a unified story of how we can come to possess different kinds of concepts. We have already seen how concepts of low level properties – paradigmatically, natural kind concepts – can be acquired ethocentrically. Also, we have touched on the possibility that ethocentrically acquired concepts can be concepts of response-dependent subject matters, and so that concepts of high-level and low-level properties can be acquired in essentially the same way.

Natural kind concepts and concepts of response-dependent subject matters are commonly treated as opposite ends of the spectrum. But there is a third option: concepts for which some variety of relativism is appropriate (henceforth, ‘relativist concepts’).²⁸ Relativist concepts differ more from natural kind concepts than response-dependent concepts do, making response-dependence of subject matter a middle position between two extremes. Unlike natural kind concepts, response-dependence of subject matter entails a form of subject-dependence, but also a moderate realism: If something is red, it is red for all humans, normal-sighted or colour blind, in good viewing conditions or not. If two people disagree about the colour of something, then at least one of them is wrong. This moderate realism goes missing for relativist concepts. For these, faultless disagreement is possible. Paradigmatic examples are matters of taste; you think marmite is delicious, I think it is disgusting, and none of us have to be wrong. We might agree to disagree, and I might even concede that the stuff is delicious to you. In relativist matters, there is no one fact of the matter – not even one that is constituted by best opinions. This makes for a higher degree of subject-dependence and a lower degree of realism than response-dependence.

How might relativist concepts come into being? And how could the relativities in them, where unrecognised, have failed to be noticed? The ethocentric story provides an attractive answer to both questions. The story about concept evolution

²⁸ Again, the options distinguished here are in-principle-options; I prefer to remain neutral on whether they are instantiated, or by which concepts.

for relativist concepts would be the same as in the case of natural kind concepts and response-dependent concepts: subjects notice a salient similarity between certain things or situations, and there is sufficient constancy in the relevant responses over time and for different subjects for a term for the common factor to be of use in communication. However, for relativist concepts the discounting practice breaks down at an earlier stage than it does for response-dependent concepts. The discounting practice does not deliver a set of C-conditions in which everyone's responses would be the same, but only delivers conditions in which there is a more limited amount of constancy: constancy in the verdicts of (time segments of) single persons and for smaller groups of people who share the same taste, sense of humour, etc. (but still enough for the concepts to serve the purpose of communication²⁹).

Relativist theses say that the (content and/or) truth value of certain propositions is relative to something – people, contexts, times, etc. – that we might not have expected them to depend on. This would sound mysterious on some views, but the ethocentric story provides an explanation of how this situation could have come about, and how the relativities could have failed to be noticed. The ethocentric story is compatible with all kinds of unarticulated constituents because all it takes for a concept to form is a salient similarity and sufficient constancy in responses for a discounting practice to work and a seems/is distinction to form. This process can 'gloss over' some relativities if their influence on the response-patterns and discounting practice is sufficiently limited or sufficiently stable. What matters in ethocentric concept formation is 1) a salient similarity – no matter how it comes about, and no matter what differences are hidden below the surface – and 2) the differences that appear in the course of the discounting practice that gives rise to the C-conditions. Any relativities that do not reveal themselves in this process will be 'overlooked' in the process of concept formation.

²⁹ This is not only a question of agreement in responses; what makes concepts of e.g. taste work despite the differences in people's responses is also agreement about the 'out rules' for the concepts – that if something is disgusting, you stay away from it and refrain from forcing it on your dinner guests, etc. The ethocentric story can easily be adapted to include such components.

This point is relevant for response-dependent concepts, too. For example, the ‘relativisation’ to standard subjects in colour concepts is easily overlooked. Presumably this happens because the contribution to colour experiences by standard subjects in standard conditions is constant, and the part of the system that varies is the object, and hence colours and colour changes get ascribed to objects, while the subject’s contribution is not recognised. This sort of explanation enables a response-dependence theorist to answer challenges of the sort ‘Explain why noone noticed the response-dependence!’.

A theory of relativist concepts – especially where the relativism is unexpected – requires exactly this feature: that the concept can form in a way that may ‘overlook’ or abbreviate away some relevant parameters. On the ethocentric story, relativities are easily accommodated, and might even be expected for cases where responses are less similar across people, contexts, etc. than we might have thought.

Relativist views come in two main versions (labels vary, but I shall use the following, from Wright (2006)). One is indexical relativism, according to which your claim ‘Rhubarb is delicious’ and mine have different content; yours says that rhubarb is delicious to your taste, mine that rhubarb is delicious to my taste. The other is ‘true relativism’, according to which your claim and mine have the same content, but different truth values. An ethocentric story is compatible with both of these versions, and is presumably neutral on which one is preferable, in general and within a given domain. This is probably a further advantage of the view; the lack of consensus about which version is appropriate for which purposes suggests that there is nothing in the way the concepts come to be possessed that determines the issue.

Extending the ethocentric story to relativist concepts is taking it very far from its origin. To Pettit’s taste, even response-dispositional concepts are too subjectivist, and his main motivation for the ethocentric story is the way it can be used with concepts that refer to objective, subject-independent properties. It also takes us away from our main topic, so I shall not pursue the matter further in this work. But if the suggestion can be made to work, the potential of the ethocentric story will be even more impressive; the ethocentric story will provide a unified story about how concepts as different as natural kind concepts, response-dependent concepts and concepts of relativist subject matters can develop, using essentially the same

materials. The story would also explain how it is that we often expect objectivity in a given domain and ultimately have to settle for less, even if the concepts do not thereby become useless or subject to error theory.

The naive view about reference for most concepts is probably that the concept unproblematically refers to a property that is in common between the samples, independently of how it appears to us. Only when we discover problems with this view, e.g. irresolvable inter-personal discrepancies, or a discovery that no well-behaved, objective property is shared by the samples, do we start to consider alternatives. So low level reference is probably the first choice in many cases. An amusing example of such objectivist expectations is a story about an 8-year old who thinks of the property of being 'cool' as an objective property that people just have more or less of, as a perfectly objective and probably unchangeable matter. Also, Pettit is presumably right in pointing out that it is tracking of low-level features – or, perhaps more accurately, tracking of low level differences – that make concepts useful for finding our way in the world. (This can be true even though usefulness does not require tracking of simple, well-behaved low level properties; colour concepts are very useful though the low level properties they track are very complex and disjunctive.)

When discoveries are made that draws this objectivity into question, response-dependence and relativism serve as fallback options. If no low level referent is available, we go for a higher level referent, and may have to retreat all the way to response-dispositions or similar (though response-dependence can also come about in a different way; see Ch. 6). But it can happen that there is not enough constancy in responses for the ethocentric story to do its work for the population as a whole, and deliver a set of C-conditions in which everybody would agree in their verdicts. Constancy may be obtainable for smaller groups, e.g. those sharing a sense of humour. In such cases, we get concepts for which some kind of relativism is appropriate.

3. How are the ethocentric stories related?

In the remainder of this chapter, we shall consider the relation between Pettit's ethocentric story of concept acquisition and the application to concept evolution explored here.

The ethocentric story of concept evolution is a very natural extension of the ethocentric story of the concept acquisition of individuals, but an extension all the same. It seems natural, but not mandatory, to combine the two accounts. One might accept one and reject the other; for example, one might hold that concepts are acquired by individuals in accordance with the ethocentric story, but that they came into being in a very different way, e.g. by being invented by a God who subsequently taught them to Eve and Adam by way of exemplars in the ethocentric way (this story is far-fetched, but still conceivable).

Conversely, one might have reasons against endorsing the community version, even if one accepts the individual version. One such reason is the lack of evidence. While we all have experience with individual's concept acquisition from ourselves and others, we don't have similar evidence about concept evolution. Therefore, the empirical basis for the latter thesis might be considered too weak, and a charge of illicit armchair science may seem a more imminent threat; even if the story is intended as an in-principle-story about how concept evolution could be possible and not a story about how concepts actually evolved, it might be thought that it has to be backed up by empirical evidence that it could be on the right track.³⁰ This might perhaps motivate the combination of acceptance of the individual version with rejection (or at least non-acceptance) of the community version. But on the whole, the two versions may be expected to come and go together.

³⁰ There are clearly questions here that need to be addressed. What is an in-principle-story? If it's a story about how something *could* have happened, what kind of 'could' is involved? It is clearly not the 'could' in 'a sorcerer could have brought language into existence'. A story of this sort is presumably only interesting if it is informed by the empirical facts we know about the subject matter, if it is consistent with the rest of our knowledge, and if it has some plausibility from our perspective. Is it legitimate to do this kind of reasoning from the armchair? This might seem questionable. On the other hand, it also seems to be the best we can do when trying to address questions like the Wittgensteinian ones that provide Pettit's starting point. This deserves further discussion, but I will leave it for another occasion.

3.1 Pettit and the ethocentric story of concept evolution

It is not clear whether Pettit himself would accept an ethocentric story of concept evolution. He presents his account as a story about the concept acquisition of individuals. But many features of the story suggest a wider scope, and would make a lot more sense if the story was about concept evolution. One such feature is the account of the discounting practice that gives rise to the C-conditions, in particular the part about inter-personal constancy. While the account can be adapted to suit the case of the concept acquisition of individuals, Pettit presents the matter in a way that suggests that he is really talking about communities. Also, the claim that ethocentric concepts may refer to substantial properties, and the explanation of how this could be, would seem more appropriate against the background of a story about concept evolution; as mentioned earlier, it is hard to see how the way subjects learn a concept should rule out its reference to substantial properties.

Finally, the account of how the ethocentric story provides an answer to the problem of rule-following would make more sense for the community version. The individual version, strictly speaking, would show only how an individual could come to grasp a rule pointed out to her by a teacher who is already part of a functioning linguistic community with a practice of following rules, and how she could gain knowledge of the C-conditions from having her use corrected by other competent users. It would thus only explain how a learner could cotton on to conceptual content that already existed. But it would be silent on the natural next question: how could conceptual content come into being in the first place? To answer this question, the community version of the story is needed. This version offers an account of how, given the response-dispositions of the potential members of a linguistic community, rule-following and a linguistic practice might come into being.³¹

³¹ The story about individuals' concept acquisition would be sufficient only on a conception of rules as platonic, pre-existing entities which have nothing to do with a communal practice. But this conception of rules seems alien to Pettit's project – first, because its inadequacy motivates the very challenge that Pettit's account is allegedly designed to address, and secondly because the ethocentric story makes it superfluous; the community version of the story can explain how conceptual content may come into being without assuming the existence of platonic rules.

The terminology Pettit uses to distinguish his account from other response-dependence accounts strongly suggests that he intends it to apply to communities as well as individuals. In a recent overview article,³² he describes his view as a thesis in explanatory semantics rather than a thesis in descriptive semantics: a thesis about how concepts *come to refer* to the properties they refer to, rather than a thesis about *what sorts of properties* they refer to.³³ This is a different question from that of individuals' concept acquisition, but it would be answered by an account of the concept acquisition of communities. At other times, Pettit presents the ethocentric story as a *genealogy* of rule-following (as opposed to an analysis of the concepts).³⁴ This strongly suggests that the ethocentric story is intended as an account of concept evolution, and not just of the learning process of individuals.

However, other formulations of Pettit's suggest something different, e.g. this:

Another thing that the account given does not do is to provide a natural history, however speculative, of how rule-following might have emerged among our kind, or in other creatures. It is one thing to try to show that certain deep problems that any natural history must face are not as daunting as they might have seemed in the light of the Kripke-Wittgenstein argument. It is quite another to develop a natural, as-if history of rule following that would show how every stage in the evolutionary and cultural advance towards the following of rules can be described – and be described at a level of convincing detail – in naturalistically unproblematic terms. (Pettit (2005), p. 7-8)

The first sentence of this passage might come across as a rejection of a priori accounts of concept evolution, and certainly a statement that his account is not meant to entail one. When we look at the whole passage, however, the moral is less clear. The point may just be that the ethocentric story does not provide *a detailed, naturalistic story* of the evolution of rule-following, but only an account of how it could be possible for a linguistic practice to get off the ground, and this much is true of an ethocentric account of concept evolution.

³² Pettit (2006).

³³ The distinction is Stalnaker's (2004).

³⁴ E.g. (1991) p. 602 and (2005) p. 8.

The way I have developed the account in this chapter relies heavily on the levels framework. The moves that do so should not be ascribed to Pettit, since he does not endorse a levels framework, and presumably does not believe in levels in the sense I am interested in.³⁵ He takes the primary interest of the ethocentric story to be that it can take us to response-independent properties, and seems to think that (what in our terms would be) high-level location amounts to admitting that there is no real property there.

All things considered, we should be careful with attributing the ethocentric account of concept evolution to Pettit. It is not the story he tells. But it is a very natural extension of his account, and it may be the story he should have told, given that it is built from the materials he offers, and seems better suited than the original version to address the challenges that motivated it. Whatever the answer to the exegetical question, it is independent of the more interesting question which has been our main focus: The philosophical mileage promised by the proposal.

3.2 Basic concepts

Finally, a few words about basic concepts. As mentioned, Pettit remains neutral on which concepts are basic for the purposes of the ethocentric concept acquisition of individuals, but stresses that they can be different for different people. In the light of the community version of the ethocentric story, we might add that the concepts that are basic for a community may differ from those that are basic for individuals. Obvious candidates for basicness in the community version (as well as the individual version) are concepts of traditional secondary qualities – colour concepts, and concepts of smells, sounds, pains and other bodily sensations. Also, basic inference patterns like *modus ponens* seem so fundamental to our thinking that it is hard to imagine how they could be acquired if not in something like the ethocentric way (i.e. primitive pattern recognition and dispositions to proceed in particular ways, given that nothing disturbs the normal course of things). It is also natural to imagine this sort of mechanism at play in very early ‘concept’ formation relevant to survival –

³⁵ Occasionally, he mentions ‘orders’ of properties in the functionalist sense (e.g. 1998, p. 116), but as we saw in Ch. 3, this is something quite different.

good to eat, bad to eat, dangerous, predator, water, etc. However, all this remains speculation. The story is intended as an in-principle-story, and pronouncing particular concepts to be basic and ethocentrically evolved would take us firmly across the boundary and into the domain of empirical science. Who knows which concepts have evolved in what way? This is definitely not a question to be settled from the armchair, even if, while comfortably seated, we are free to consider how such a thing as concept evolution could be possible.

4. Summary

Pettit's version of response-dependence offers very interesting prospects. It provides an interesting response to a version of the Wittgensteinian rule-following problem – 'how can a finite set of examples enable a subject to grasp a rule with a potentially infinite range of applications?' – though it has been impossible to do justice to this issue here. It also contains the materials for a promising account of the genealogy of concepts, and whether or not an ethocentric story of concept evolution was intended by Pettit, the story provides many useful results. In particular, it provides answers to our three 'hard questions'. Let us sum up the answers that have emerged from our discussion.

- How can concepts come into existence, be used, and be in good standing before the exact nature of the referent, or even its level, is known?

All it takes for a concept to form is a salient similarity and sufficient constancy in responses for a stable set of C-conditions to form and provide a seems/is-distinction. This leaves open the question of the location of the referent. It could be a low level property; such properties can have their say in determining concepts' extensions, and can serve as their referents. Or, if no appropriate low level referent exists, or if the concept works in such a way as to make high level location preferable, it can be located on a high level. None of these options are either precluded or preferred by the core components of the ethocentric story.

- How can concepts made of 'for us'-material come to refer to independently existing properties and kinds?

For essentially the same reasons as above; the ethocentric story allows that a concept developed on the basis of a response-pattern can refer to a property that is

independent of this response-pattern, e.g. the low level property correlated with responses in C-conditions, and causally responsible for these. So on the basis of our response-dispositions, we can acquire concepts that refer to response-independent properties, and yet be confident that our applications of the concepts are often correct (since a property only gets to be the referent of the concept if it is (roughly) correlated with responses in C-conditions; if there is no such property, then higher level locations provide a fallback option).

- Can we tell a unified story about how different kinds of concepts – natural kind concepts, concepts of response-dependent subject matters, and relativist concepts – come into being?

Yes, the ethocentric story of concept evolution provides such a unified story. The recipe of salient similarity patterns and a discounting practice aimed at identifying perturbing and favourable factors and seeking as much constancy as possible fits concepts of all three sorts, depending on the existence of appropriate low-level reference candidates, and the amount of constancy feasible for the subject matter in question before the discounting practice breaks down.

The account has a further advantage that we shall pay more attention to in Ch. 6: It offers a useful way to distinguish between different routes to a response-dependence thesis of subject matter.

The account of concept evolution suggested in this chapter is just a rough outline and needs further work, and objections will almost certainly be forthcoming. However, it offers very interesting prospects if it can be made to work. In this chapter, I have focussed on the advantages of the proposal in keeping with the aim of making sense of response-dependence.

Ch. 6: Response-dependence of subject matter

This chapter, like Ch. 4, is mainly concerned with response-dependence theses of subject matter. The main questions to be considered are these:

- how should we formulate (i.e. capture the intuitive content of) the distinction between response-dependent and –independent subject matters? and
- how can the distinction be made operational?

Though I offer no final verdict on how a response-dependence thesis of subject matter should best be formulated, it should come as no surprise that I suggest a level-based way of thinking about the distinction: response-dependence theses are theses of high level location. In the first part of the chapter, I describe two routes by which we might arrive at a conclusion of response-dependence of subject matter. One is an a posteriori route via low-level reference failure of level-flexible concepts. The other is an a priori route via transparency intuitions etc. to the conclusion that the concept is high-level rigid and criterially governed. I discuss how these routes combine with the versions of response-dependence from the literature.

The second part of the chapter concerns two points of disagreement between Johnston and Wright: the question whether response-dependence theses should be formulated in terms of dispositions or subjunctive conditionals, and the question whether response-dependent concepts are criterially governed concepts or referentially governed concepts that refer to response-dispositions. I argue that there are no big decisions to be made regarding these matters, as both disagreements make less of a difference than normally assumed. A brief introduction to the location dispute for dispositions forms the backdrop of this discussion.

The third and largest part of the chapter concerns the question how to make the distinctions operational, i.e. helpful in determining the response-dependence status of a domain about which we are in doubt. I examine two of the conditions normally thought useful for this purpose: the a priority and necessity conditions (most of the other relevant parameters have been or will be discussed elsewhere in the thesis). I argue that to the extent that these conditions will serve the purpose, they will do it by means which can do the job independently of the necessity/a priority conditions. I

suggest that in order to determine whether a domain is response-dependent, we should also take into consideration a broader range of features that are normally considered in location disputes, such as transparency intuitions and similar. The criteria we settle on do not guarantee a clear-cut judgement for every concept, however, as many everyday concepts come with conflicting intuitions.

1. Response-dependence theses as high level locations

In the light of the level-framework, response-dependence theses like Johnston's and Wright's – response-dependence theses of subject matter – can be understood as high level locations. This way of viewing them does, in my opinion, offer a fruitful perspective on response-dependence of subject matter. It also offers a way of describing what is common ground between all the versions that aim to capture some sort of response-dependence of subject matter – i.e. all versions except Pettit's (and Haukioja's Pettit-style view; as noted, such views are neutral on the location issue, though Pettit prefers what in our terms would be low level locations). The relevant high level location is described differently by different authors: To Johnston and others, the relevant high-level properties are response-dispositions, to Wedgewood they are properties with response-dependent essences, and to Wright, they are the deflationary properties corresponding to concepts that have their extensions determined by best opinions.

There are theses of high level location that are not response-dependence theses, e.g. (role property) functionalism and projectivist views. So saying 'high level locations' is not enough; in order to capture the class of views we want, more needs to be said. Wright's terminology of extension-determining seems a fairly neutral way of saying this: response-dependent properties are properties that have their extensions determined by responses in C-conditions.

This very general description leaves a lot of open questions about the intuitive content that response-dependence is intended to capture, and how to make the distinction operational (indeed, that's why it will serve as a general description). For example, it leaves open which of the above-mentioned ways of capturing the distinctions is the better one. I shall argue that some of the differences are less significant than they might seem. However, there is another (and in my opinion more

important) distinction which is normally overlooked, and which is brought into focus by the levels framework and the discussion from Ch. 5. It concerns two different ways response-dependence theses might be reached, rather than different things they might be saying. This distinction is the topic of the next sections.

2. Two routes to response-dependence of subject matter

The results from Ch. 5 put us in a position to distinguish two routes by which a conclusion of response-dependence of subject matter could be reached. (In this chapter, ‘response-dependence’ means response-dependence of subject matter unless otherwise stated.) One is by a priori considerations alone, while the other is an a posteriori route that involves empirical results in the settling of the location question. The a posteriori route is relevant when a level-flexible concept turns out to lack a suitable low-level referent, and is ultimately located on a high level as a result of empirical investigation into underlying properties. This route is suggested by the ethocentric story told in Ch. 5 of how a concept can come into existence and be used before the exact nature of the referent, or even its level, is known. The idea would be that, in keeping with a natural tendency towards objectivism (cf. the young objectivist about coolness), we expect a completely objective (low-level) referent for the level-flexible concept. When faced with the lack of a suitable low-level referent, we turn to higher levels in search of a referent. In some cases, the best candidate turns out to be a response-disposition or similar, and we might then conclude that this is the referent of the concept. Response-dependence in this sense would be an a posteriori conclusion. Johnston’s (1992) case for a revisionary response-dependence thesis about colours is a very clear example of this route to response-dependence, since it puts a lot of weight on empirical facts, most notably the lack of a (non-disjunctive, unified) physical property common to all and only yellow things.

The second route to response-dependence is an a priori route. It is relevant where it is clear from the way the concept works that it could refer to nothing but a response-dependent property. The evidence for such a conclusion could be strong and durable intuitions about transparency or similar, or intuitions that responses in C-conditions are the highest court of appeal, whatever the situation turns out to be on lower levels. (These and other relevant criteria will be discussed later in this chapter.)

The difference between these two routes to response-dependence could be formulated in terms of the distinction between sense and reference.¹ If the *sense* of the concept makes it clear that it would follow responses in C-conditions in all scenarios, then the concept can be judged response-dependent a priori. If, on the other hand, the (sense of the) concept is such as to make it appropriate to look for a *referent*, and if this referent could be something different from a response-disposition – e.g. a physical common factor – then the conclusion of response-dependence, if reached at all, will be reached a posteriori when it turns out that the response-disposition is in fact the best reference candidate.

The two routes to response-dependence of subject matter are, once again, alternatives that can be distinguished clearly in principle, whereas for a given concept it may be hard to decide which of the stories fits best. There may be conflicting intuitions, some of which suggest high-level rigidity (e.g. transparency intuitions), while others suggest level-flexibility combined with low-level reference failure (e.g. some degree of expectation of low level reference, and empirical discoveries of low level heterogeneity leading to the realisation that the concept must be response-dependent). This may well be the situation for colour concepts. This mix of intuitions might explain how some (e.g. Wright) could think of colours as high-level rigid while others (including Johnston) clearly think of them as level-flexible. Such mixed pictures will be natural given the story told in Ch. 2 about level confluences in location disputes, and given the often mixed and manipulable intuitions about location that people have.

Does the route by which the conclusion of response-dependence is reached have any implications for what sort of response-dependence thesis is appropriate? It would seem natural to associate the a posteriori route with Johnstonian response-dispositionalism. Not only does Johnston capture the a posteriori route in his description of revisionary response-dependence theses. More importantly, concepts that turn out to be response-dependent by this route are meant to be associated with a referent with some degree of substantiality, not only a set of criteria and a deflationary property. A response-disposition, even if it is a high level property, has

¹ Thanks to Crispin Wright for suggesting this way of putting the matter, and for helpful discussion of the issue.

some degree of substantiality (there are objective facts about how people are disposed to respond, and these facts make for a classification that has some sort of independence), whereas Wright-style response-dependence will associate the concept mainly with an extension-determining criterion, and no substantial referent ever enters the picture.

This, then, is a natural way of viewing things: We have distinguished three versions of response-dependence, corresponding roughly to the three main versions familiar from the literature: First, a ‘weak’ or ‘genealogical’ sort of response-dependence that is neutral on the location issue and concerns only the way concepts are acquired or formed (though it does rule out low-level rigidity). This is Pettit’s sort of response-dependence, though the levels-based elaborations made in Ch. 5 are my responsibility. Secondly, there are two kinds of response-dependence of subject matter, or high level response-dependence: First, there is one that comes about as an a posteriori conclusion when a level-flexible concept fails of low level reference. This route is naturally associated with Johnstonian response-dependence, where the referent of the concept is a response-disposition. In particular, it fits well with Johnston’s account of revisionary response-dependence theses. The third version of response-dependence can be recognised to obtain a priori, from considerations about transparency intuitions etc. This version is naturally associated with Wright-style response-dependence and the view that the concepts are criterially governed. Both of the latter versions are theses of high level location.

2.1 Fitting in Johnston’s contrast between descriptive and revisionary versions

Like most nice and simple pictures, this nice and simple picture over-simplifies things. First, the fit between our three routes to response-dependence and the three versions from the literature is not perfect. No provision is made for descriptive Johnstonian response-dependence. This is supposed to be an a priori conclusion, but it is supposed to go with a dispositional response-dependence thesis which entails referentiality as much as the revisionary version does. If he is right, we should not associate referentiality solely with the a posteriori route and criteriality with the a priori one. (As I shall argue, this is no serious problem, as the criteriality /referentiality issue turns out to make little difference in these cases.)

Even if we bracket this issue, Johnston's distinction between descriptive and revisionary response-dependence theses does not coincide exactly with the distinction between the a priori and a posteriori routes. A revisionary response-dependence thesis might come about because a low-level common factor is found to be missing for a concept that is apt to survive this discovery. But it might also come about a priori as a result of thinking through the intuitions connected with the concept, perhaps finding, to our surprise, that the concept is a high-level rigid one. For example, we might have thought that pain was apt for location with a low level physical property, and then learn from Putnam's and Kripke's arguments that the concept is in fact high-level rigid.² Or it might turn out (again, by a priori reasoning) that the intuitions that come with the concept pull in different directions and are strictly speaking inconsistent, and that the best way to purify the concept of the inconsistencies is to think of it as a high-level rigid one. This would be an a priori conclusion, but it might still constitute a revision of the concept or the way we used to think about it. In other words, the distinction between descriptive and revisionary response-dependence theses may well be a question of familiarity versus surprise, rather than of empirical versus a priori routes to the conclusion.

2.2 Different routes; different theses?

The simple picture of the a posteriori route leading to Johnstonian response-dependence and the a priori route leading to Wright-style response-dependence is questionable for more interesting reasons than the exegetical matters discussed above. Two seemingly important points of disagreement between Johnston and Wright are 1) whether response-dependence should be formulated in dispositional terms or in terms of subjunctive conditionals, and 2) whether response-dependent concepts are referentially governed with response-dispositions as referents or criterially governed with responses in C-conditions as the main criterion. (These two questions may be related, but can in principle be kept apart, so we should treat them as separate questions.) I shall argue in the following that the differences in both

² The arguments I have in mind are the multiple realisation argument (e.g. Putnam 1967) and Kripke's argument against identity theories of the mind (1972, p. 144-55).

respects are a lot less significant than they seem; they are merely results of different ways of looking at the same matters. To see how, we need a look at the location dispute for dispositions, as the answers to both questions depend heavily on what view is taken on the location question for dispositions. This involves some extra stage-setting, but it should help us answer these two crucial questions about how response-dependence theses of subject matter should be formulated.

3. The location dispute for dispositions

Dispositional concepts – concepts like solubility, fragility, etc. – give rise to a location dispute similar to the one about colours outlined in Ch. 2. (There are dispositions that do not fit into this picture, e.g. divisibility by 2, but we can ignore this complication here.) As with colours, different levels of properties are relevant to dispositions. First, there are the low-level properties that serve as bases of the dispositions, i.e. do the causal work associated with the disposition. For solubility, for example, the base property will be certain bonds between molecules which will change and release the molecules on immersion in a liquid. The base property can itself be of either categorical or dispositional nature; if the latter, it will often (always, in the opinion of many) be possible to find bases for that disposition on even deeper levels. For example, the base for the disposition to elicit colour experiences in humans with normal colour vision is the disposition to reflect light with certain distributions of wave-lengths; the base for this disposition, in turn, is surface structures or similar. Perhaps there are even deeper levels of base properties that we don't yet know about.

On another, higher level of relevant properties, we find the patterns of stimuli and responses characteristic of the dispositions in question – e.g. dissolving when immersed in liquids, or eliciting colour experiences in suitable observers in C-conditions. The classical conditional analysis of dispositions treats these patterns as essential to dispositions. It can be stated as follows:

- (1) Something x has the disposition D to elicit a certain response r to stimulus s in locus l iff x would elicit r in l on s ³

This response pattern and the base of the disposition (and perhaps the ultimate, lowest-level base of the disposition, supposing there is such a property) constitute rival location options in the location dispute about dispositions. They are correlated in the actual world (or, more precisely, in the subset of circumstances in the actual world that count as relevant to the disposition in question), but they can come apart in principle, and often in practice (even if the bases are (actually, normally) causally responsible for the response-patterns). The usual questions arise: Which of the candidates should the dispositional concepts follow? Which aspect is most essential to the disposition?

The whole range of standard arguments are available. First, there are arguments from causal efficacy for a low level location. Dispositions can act as causes; an animal's disposition to learn from its mistakes causes it to survive, the dullness of a scalpel causes an operation to go wrong. But the causal work dispositions do is done by their bases. So, the argument goes, in order to avoid causal over-determination, we should conclude that dispositions are identical with their bases.

(Another argument for a low-level location – or at least for the relevance of low-level features to dispositions – might be found in the consideration that dispositions can be manipulated by manipulating underlying physical properties. Examples are omnipresent – e.g. firing ceramic vases in order to make them waterproof and less fragile, taking anti-depressants to remove or mask the disposition to excessive emotional responses to everyday stimuli, etc. This reasoning is completely parallel to similar arguments for the relevance of states of brains to states

³ This is not quite the traditional formulation, as it has been adapted to accommodate dispositions to elicit responses in other things and beings as well as dispositions of x itself to respond in certain ways (the special cases where $l=x$); normally, only the latter case is taken into account. In Ch. 8, we shall discuss some problems with the conditional analysis of dispositions that it shares with response-dependence theses; for now, we shall stick with this simple formulation.

of minds. It does not support an identification with the low-level properties, though, but merely a conclusion of fairly strong supervenience.)

Likewise, the usual argument types for high level locations are available. There are arguments from unity: In many cases – fragility, poisonousness, etc. – one and the same disposition has different bases in different instances. But we wouldn't like to say that e.g. fragility is a different disposition for eggs, china vases and thin icicles; intuitively, they share the property of being fragile in virtue of their tendency to break. So dispositions are distinct from their bases. (Or, if they were to be thought of as low-level properties, they would have to be disjunctions of bases, not bases simpliciter. But this would make them as unfit to be causes as higher-level properties are, and so undermine the causal argument. This line of thought is found in Johnston (1992) for the case of colours.)

There are also arguments of the type we termed transparency arguments in the case of colours and mental states (though perhaps the label is less fitting for these cases). An example: 'If something elicits nausea, there is no further question whether it is nauseating. But the physical properties that cause the unpleasant reaction might not have had this effect, and so these can't be the nauseatingness itself.'

Perhaps the most important argument in the location dispute for dispositions would be an argument from pragmatic considerations. Dispositions and disposition ascriptions matter because it is extremely important to be able to form expectations and make predictions of what will happen in various circumstances, and what will happen if you take certain actions and manipulate your environment in certain ways. To ascribe a disposition to something is to say that it can be expected to respond in certain ways in certain situations. This is what gives a conditional analysis of dispositions its appeal. For pragmatic purposes, what matters is performance. Knowledge of bases may matter too, of course, but mainly (only?) because it is an efficient tool for manipulating and predicting performance. Knowledge of low level facts can be helpful for these purposes, but you can also manipulate and predict performance without knowledge of bases. This aspect is captured by a high-level account of dispositions, but not by a low-level one.

All the possible positions described in Ch. 2 with the example of colours are available in the location question for dispositions. We can choose a location, higher

or lower (or intermediate, if we locate with a base property that is itself of dispositional nature). Or we can recognise that the concepts equivocate, and leave the matter there. Or we could argue for a multi-level location, e.g. with role properties ('the property of having a property that brings about the characteristic responses as a result of the stimuli' or similar) and their realisers (the bases that do this causal work). Finally, we could be eliminativists and give up the notion of dispositions altogether, or we could be primitivists and conclude that dispositions are *sui generis* phenomena that are not to be analysed in terms of stimuli and responses or in any other way.⁴

I favour a high level location of dispositions. I also think the conditional analysis goes with a high-level location, and that we won't get anything better than this (even if the simple conditional analysis has to be supplemented with a 'we know what we mean'-component in the form of a *ceteris paribus* clause, C-conditions, or a normality-based semantics for counterfactuals in order to avoid conditional fallacy problems; we shall discuss these matters further in Ch. 8). But I won't be able to take a well-argued stance on the location dispute for dispositions in this work; a proper survey and evaluation of the arguments in the location dispute about dispositions would take up several chapters in itself. Instead, we shall look at the consequences of different location choices for response-dependence accounts, specifically with respect to the two questions mentioned above: subjunctive versus dispositional formulations, and criteriality versus referentiality.

4. Dispositional or subjunctive formulations?

The answer to the location question for dispositions has strong implications for the relationship between subjunctive and dispositional formulations of response-dependence theses. If dispositions are located on a high level, with response-patterns, the two will not be very different. In particular, if the high level location is cashed out in terms of a simple conditional analysis of dispositions, then they will be equivalent, as dispositional formulations of response-dependence theses will be analysable in terms of their subjunctive 'competitors'. It may be that a simple

⁴ E.g. Martin (1994).

conditional analysis needs to be modified in response to the problems to be discussed in Ch. 8, but if so, the subjunctive claims will presumably need the same kind of modification for the same reasons, so this shouldn't make a difference.

If, on the other hand, dispositions are located with their bases (or ultimate bases), then the two kinds of account will come out very different. The subjunctive formulations will still go with high level locations, while the dispositional version will go with a low level location. If a low-level location is chosen, it might be questioned how much response-dependence is left in the proposal. The resulting thesis won't be response-dependence of subject matter; the base property will normally exist independently of the responses in all interesting ways, and have its extension determined independently of the responses it happens to elicit. At most, it will be response-dependent in the (reference-fixing) sense that this property gets to be the referent of the concept because it is the property that elicits the relevant responses in the relevant subjects and conditions. Response-dependence claims combined with such a location would surely vindicate realism. But the subject-relatedness would be a lot thinner than we might have expected.⁵

A low level location would be incompatible not only with my way of formulating response-dependence of subject matter explicitly in terms of high level location, but also with most other views in the literature, including Wedgwood's suggestion about response-dependent essences and Wright's about extensions being determined by best opinions, not by tracking extensions of response-independent properties. Also, the combination of response-dependence with low-level location of dispositions would jeopardise the necessity of the response-dependence biconditionals, as response-patterns could come apart from the base property. This would be bad news for those who endorse a necessity condition (e.g. Johnston).

How about the alternatives that are not straightforward locations? Eliminativism about dispositions is clearly uninteresting as a basis for response-dependence theses. On this view, subjunctive conditionals would be the only option left for response-dependence theories. Primitivism about dispositions would also

⁵ Considerations similar to these motivate Lopez de Sa's claim that rigidified response-dependence claims do not really amount to response-dependence, though he does not put it in terms of levels. Lopez de Sa 2003.

make the accounts different; on such a view, a response-dependence thesis formulated in terms of dispositions would allow for no further explication, while a formulation in terms of subjunctive conditionals might offer a more informative account. Equivocation views would also make the accounts slightly different, as the low level component of dispositions would presumably not be included on subjunctive accounts. If a multi-level location approach is taken, e.g. a functionalist thesis, the situation would presumably be the same; a dispositional and a subjunctive formulation will be similar to whatever extent the weight is on a high level location (e.g. with the role properties of a functionalist account), and different to whatever extent the emphasis is on bases.

Given my preference for a high level location of dispositions, and for some sort of conditional analysis as the best way of describing them, I don't think that there is a big issue regarding the choice between dispositional and subjunctive formulations. A high level location will make the two types of account equivalent, provided that dispositions can be analysed in terms of subjunctive conditionals, which I expect that they can, either by way of a simple conditional analysis or by some modification of it.⁶ A convincing case for this conclusion would, of course, involve showing that a high level location and an analysis in terms of subjunctive conditionals is the right approach to dispositions. But this is a task I cannot take on here.

5. Substantial referents or criterially governed concepts?

Wright-style response-dependence can be paraphrased as the view that the concepts in question are criterially governed with responses in C-conditions as the dominant criterion – or, more accurately, with responses of the relevant kind in the absence of known perturbors serving as a defeasible criterion, defeasible by evidence that the C-conditions are not met. No substantial referent ever enters the picture on this view, though there may be substantial properties that are coextensive with the criterially governed concepts. Their referents, if any, will be purely deflationary properties. On Johnston's view, by contrast, response-dependent concepts are concepts of response-

⁶ See Ch. 8 for discussion of some possible amendments to accommodate conditional fallacy problems.

dispositions, and thus have referents of sorts, even if these are not low-level, subject-independent properties. This would seem to constitute an important difference between the two views.

Criterially and referentially governed concepts are normally thought to function in very different ways. Criterially governed concepts have their extension determined as whatever fulfils the criteria, whereas with referentially governed concepts, the extension-determining job is deferred to a (more or less substantial) property that has an extension in its own right, and this extension determines the extension of the concept. The two are normally thought to be incompatible, as there can be no a priori guarantee that the extension picked out by a criterion coincides completely with that of a substantial property.

Wright has objected (in discussion) that the dispositional version of response-dependence obscures the clear distinction between referentially and criterially governed concepts by talking about response-dispositions as if they were properties substantial enough to serve as referents – a sort of intermediate stage between ‘real’ natural properties and the purely deflationary properties that Wright associates with response-dependence. He thinks such an intermediate position is a misunderstanding. I shall argue, however, that with respect to response-dependent concepts, the contrast between criteriality and referentiality is not the crucial divide it might seem to be (and would be in other areas).

5.1 Two faces of response-dispositions

I take the two approaches to be compatible for the following reason: There are two ways of viewing response-dispositions, which could be described as a first person and a third person perspective. Viewed ‘from the outside’, a response-disposition is an objectively existing pattern of similarity in reactions across a range of subjects, situations, and ‘objects’ (with ‘objects’ understood in a sufficiently wide sense, and still presupposing a high-level location of dispositions). There are objective facts about how people are disposed to judge, see, feel nauseated etc. These patterns may be taken to constitute properties that have some degree of substantiality and are capable of determining extensions of concepts, and of serving as referents of referentially governed concepts.

But when response-dispositions are viewed from the perspective of one of the responding subjects, things look very different: response-dispositional concepts apply whenever the criteria – reaction to stimulus in circumstances – are fulfilled. There is nothing more than this to guide the extension of the concepts. There is no property that you're trying to track in your responses, not even in the way there is when you're trying to keep in step with the cool people about matters of consensus. What happens on lower levels may be causally relevant, but what determines the extension is the response pattern on the high level. The low-level story is in a sense irrelevant to this.

This Janus-faced nature of dispositions would explain how Johnston's and Wright's views can be as similar as they are, even though Johnston treats response-dependent concepts as referentially governed, while Wright treats them as criterially governed. They simply focus on different 'faces' of the response-dispositions; Johnston's account approaches them from the third person perspective, whereas Wright's (especially his account of intentional states) focuses mainly on the first person perspective. This is probably why response-dependent concepts look criterially governed to Wright, and referentially governed to Johnston.

It also suggests that the two routes to response-dependence distinguished in the beginning of this chapter might lead to very similar theories, though the conclusion of response-dependence comes about in very different ways on the two accounts. The a priori route focuses on the first person perspective, while the a posteriori route takes the objectified response-pattern as fallback referent when nothing better (more objective) can be obtained. (There are other differences, though; more on these at the end of this chapter.)

At the risk of over-simplifying, we might say that the first person perspective delivers the story about the assertibility conditions of response-dependent concepts: 'apply the concept whenever it seems right (you have the experience in question, or whatever) and there are no known perturbers, but accept that your judgement may be overruled if it turns out that the C-conditions were not met'.⁷ This would explain our

⁷ This would work most straightforwardly if the relevant responses are phenomenal responses. But as Wright has pointed out (in discussion, summarised in Ch. 4), there are presumably response-

epistemic access to the subject matter in question – the fact that we often get it right, and sometimes wrong. The third person perspective provides truth conditions, and an objectively existing, though subject-related, property in the world for assertions to be measured against: the response-pattern. This would go some way towards establishing a modest realism about the subject matter in question.

Distinguishing the two perspectives enables us to answer an objection of Blackburn's to response-dependence theses:⁸ If to be F is to be such as to elicit R from S in C (or be disposed to do so, or whatever formulation you prefer), why is it that I'm allowed to say e.g. that a play was boring just because I was bored when seeing it, and without investigating how everyone else would react? And why is it inappropriate to say that the play is good if I know that standard observers in C-conditions like it, but haven't seen it myself? The answer suggested by the distinction just drawn is that assertibility conditions of a response-dependent concepts go with the first person perspective, not with the third person one. Primarily, you should use the concept when you have the relevant response and no reason to assume failure of C-conditions (though provision would have to be made for correct ways to use the concept outside C-conditions, e.g. judging colours in deviant light, the influence of which you know how to correct for).⁹

To sum up the results so far, Johnstonian and Wright-style response-dependence theses are not nearly as far from each other as they might have seemed. The disagreement over referentiality versus criteriality can be viewed as a result of viewing the same matter from two different perspectives, a first person and a third person one. The disagreement regarding dispositional and subjunctive formulations

dependent concepts for which there is no phenomenal common factor. For judgement-dependent concepts, the story about assertibility conditions would have to be more complex one about a salient similarity between the cases, even though it is not a phenomenal common factor that makes the similarity salient.

⁸ Blackburn (1993), p. 274-75.

⁹ A related but slightly different answer to Blackburn's challenge would be that some concepts, e.g. the aesthetic ones in Blackburn's examples, equivocate in their use between an anthropological sense (which corresponds to the third person perspective) and an individual sense in which you should go by your own responses (which corresponds to the first person perspective).

of the biconditionals also turns out to be insubstantial, at least if we assume a high level location of dispositions, and the plausibility of a conditional analysis of some sort. The important things are common ground: a high-level location of the disputed phenomena with response-patterns (which provide both criteria and referents).

In the remainder of this chapter, I shall discuss how to make this intuitive distinction operational. The next sections focus on some problems that are common to Wright's and Johnston's views: the problems with using the 'traditional' tools of a priori and necessary biconditionals for making the response-dependence distinction operational. I shall argue that these are less efficient than normally assumed, and suggest an alternative approach.

6. Biconditionals and operationality

It is common ground between Wright, Johnston, and most other authors that response-dependence theses should be cast in terms of an appropriate biconditional or provisoed biconditional which must fulfil a range of conditions, including substantiality, a priority, and sometimes necessity, independence and the extremal condition. This combination of equations and conditions is thought to constitute a litmus test which helps us distinguish response-dependent from response-independent concepts and subjects matters. My aim in the following is to question this picture. I shall argue that the apparatus is less suited for the job than normally assumed, and that we might fare better by looking at other parameters when trying to determine the response-dependence status of a given domain.

As we saw in Ch. 1, response-dependence biconditionals come with a lot of scope for variation: there are choices to make regarding the specification of the place-holders R, S and C (rigidified or flexible, normal, statistically typical, idealised etc.), regarding the shape and formulation of the biconditional (provisional versus basic equations, dispositional versus subjunctive formulations etc.), and regarding the requirements (a priority, necessity etc.) that the biconditional must fulfil in order to count as a sign of response-dependence.¹⁰ The different ways to fill out these

¹⁰ There is also scope for variation regarding the worlds that count: they can be the closest (or all) worlds in which C-conditions obtain, or, for Frank Hindriks's (2005) 'acceptance-dependent concepts', the actual world.

parameters make for very different theses. A lot of this scope for variation is innocent with respect to the charge I am going to make. It just allows a wide range of different theses to be considered under the heading of response-dependence, and this can be an advantage as well as a challenge, though it makes it important to distinguish the different claims that might be made by way of response-dependence biconditionals.

But there is also scope for variation where this is less welcome. First, the biconditionals can be used to make very different claims in ways that are not always clearly distinguished in the debate (see next section). Secondly, it looks like the core conditions used to distinguish between response-dependent and –independent domains – the a priority and necessity conditions – can be fulfilled for very different reasons by different sorts of theses. This makes the apparatus less suited to capture the distinctions it was designed to capture.

6.1 Four readings of response-dependence biconditionals

Equations of the form

- (2) $x \text{ is } F \leftrightarrow x \text{ would (}/\text{is disposed to) elicit response } R \text{ from subjects } S \text{ in conditions } C$

can be taken to make at least four different sorts of claims.

First, the biconditionals can be read as claims about *a priori equivalencies between concepts*. Johnston (1993) presents response-dependence claims as being of this sort.

Secondly, the biconditionals can be used to state *what determines the reference* of the concept F: ‘the property that the concept ‘F’ refers to is picked out like this...’. This is the intended reading of claims of Pettit-style response-dependence (though the ethocentric story is presumably also compatible with the first and third readings). It is also compatible with some Johnstonian response-dependence theses.

Thirdly, the biconditionals can be used to state the *criteria* characteristic of a criterially governed concept: ‘this is how you sort instances into Fs and not-Fs’. This would be the relevant reading for Wright-style response-dependence.

Finally, the biconditionals can be used for making *property identity claims*: ‘Property 1 is the same as property 2; the two property names share the same referent’. Johnston attributes such claims (claims like ‘being red = the disposition to look red to standard subjects under standard conditions’) to Locke.¹¹

The possibility of employing the biconditionals to make such very different claims is not in itself a strong objection to the biconditionals, as the further conditions etc. might make clear what claim is made. But it suggests that we should be careful to settle what version is at play in each case; the biconditionals by themselves do not settle this matter. The situation is more serious with respect to the (appearance that there are) different ways of meeting the a priori and necessity conditions. This will be the subject of the following sections.

7. The necessity condition

The necessity of appropriate response-dependence biconditionals might look like a good way of separating response-dependent concepts or properties from response-independent ones. If responses in C-conditions determine the extension of a concept, or if the concept refers to the disposition to elicit R in S in C, then you’d expect the extension of the concept, and the corresponding property, to follow responses in C-conditions across possible worlds. If, by contrast, the concept refers to a response-independent property, this property can presumably come apart from the response-pattern in other possible worlds, even if they are perfectly correlated in the actual world. So isn’t the necessity of the appropriate equations the perfect test for response-dependence?

Unfortunately, it’s not that simple. First of all, a lot depends on the exact formulation of the biconditional. If the specification of the C-conditions or responses or subjects is not exactly the right one for the concept in question, the biconditional may come out contingent or false. What we should ask is really if *a* biconditional with substantial specifications of R, S and C will be necessary, or whether such a biconditional, if true, will also be necessary.

¹¹ Johnston 1998, p. 9. Such claims are necessary if true if the names are rigid designators, but it is hard to see how they could be a priori.

But even the fact that such a biconditional would be necessary for some appropriately chosen specification of R, S and C does not suffice to tell response-dependent from response-independent concepts. After all, even natural kind concepts can make for necessarily true biconditionals of something like the usual shape, e.g.

- (3) X is water \leftrightarrow x is the (stuff dominantly responsible for the co-instantiation of the features of being a) tasteless, colourless, odourless, thirst-quenching liquid, found in rivers, lakes and oceans etc. *in the actual world*¹²

What tells this case apart from the response-dependence cases (apart from the complex symptoms, which are hardly the issue here, and the ‘dominantly’ that we shall come back to in the discussion of a priority below)? The obvious answer is that the necessity of this biconditional is obtained by rigidification alone: it is the ‘in the actual world’ that does the trick. The necessity is due to the fact that the reference of the concept of water is fixed as the property (if any) that satisfies this description in the actual world, and the term is supposed to be a rigid designator of that property, which means that ‘water’ will refer in all worlds to the property doing this job in the actual world. For concepts of response-dependent subject matters, by contrast, the necessity is supposed to come about in a different way – not as a result of rigidification, but because the response-pattern determines the extension of the concept, and hence the extension will follow the response-pattern across worlds, even in worlds where the low level story about how it comes about is completely different. Is there a way to distinguish between these two routes to necessary biconditionals, and isolate cases of the latter kind?

Necessity without rigidification can presumably be obtained only for response-dependent concepts. So an obvious suggestion would be that where rigidification is *not* involved, necessity of the biconditionals will be a sign of response-dependence. However, this would not capture the class of response-dependent concepts. For some response-dependent concepts, rigidification in the specifications of R, S and C is intuitively in place, and hence no non-rigidified specifications will yield necessarily

¹² In the context of response-dependence debates, this point has been made e.g. by Wedgwood (1998) and Haukioja (2001).

true biconditionals. Wright explains the need for rigidification in the following way, taking colour concepts as example:¹³ When trying to capture conditions favourable for e.g. colour judgements, we need a specification that meets the substantiality requirement. This requirement would be met by specifications in basic physical and physiological terms. However, such a specification would jeopardise the necessity of the equations, as the physical and physiological story might have been different. A purely statistical specification ('subjects with *statistically typical* colour vision') will lead to the same problem, as e.g. dichromate vision might have been typical among humans. A rigidified specification seems to solve the problem. If we specify the subjects and C-conditions as 'humans with *actually* statistically typical colour vision in *actually* typical lighting conditions etc.', we fix the physical story, but in an indirect way that would get things right even if the physical story was a different one.

Besides these considerations about the substantiality requirement, there are others: We want to say that the colours could remain the same even if we all became colour blind, or if standard lighting conditions radically changed, say due to some environmental disaster. Rigidification seems to be the way to respect such intuitions by fixing on the responses and conditions that are actually treated as relevant.

However, as Wright notes, rigidification 'is apt to change modal status',¹⁴ as we saw in (3) above. If for some supposedly response-dependent concepts, the specifications of R, S and C include rigidification, how do we know that the necessity of the resulting equations is a result of genuine response-dependence, and not an artefact of rigidification like the necessity of (3)?

An initial response might be that the rigidifier appears in different places in the two cases. In (3), the rigidifier is up front as the main operator, and turns the entire equation into a reference fixing description. But in versions of (2) that include rigidifiers, these appear more locally in the specification of R, S and/or C ('humans with *actually* statistically typical visual apparatus' etc.). And do we have any reason to assume that rigidification on anything but the whole equation will yield necessity, except as a result of response-dependence.

¹³ Wright (1992), p. 113-17.

¹⁴ Wright 1992, p. 115.

Wright argues that unfortunately we do; rigidification on R, S and C *can* yield necessity for the wrong reasons.¹⁵ When we stipulate that the subjects and conditions are as the actually typical ones, we export a lot to the world under consideration: the visual apparatus of the observers, the detail of the conditions (light, distance, etc.), the laws of nature, etc. And if this much is bound to hold in any world under consideration, it is also likely that the results in terms of keeping in touch with the extension of the relevant property will be the same, whether or not that property is response-dependent. This might give us necessary biconditionals even for response-independent concepts. So the necessity of the biconditionals does not look like the way to tell response-dependent concepts from response-independent ones.

In response to this problem, Wright lists three strategies. One of them is the line he takes in *Truth and Objectivity*:¹⁶ to replace the necessity condition with an a priority condition, which might fare better in helping to separate response-dependent from response-independent domains. Ironically, this condition has been subject to similar objections: that it can be met even by response-independent concepts, though for different reasons (we shall return to this below).

The second strategy is best explained against the backdrop of Dan López de Sa's views on response-dependence.¹⁷ He argues that the issue of rigidification marks an important distinction between two forms of response-dependence: rigid and flexible response-dependence. For the former, the biconditionals contain rigidifiers, and allow the concepts to latch on to real, if dispositional, properties (this is the case for colours and secondary qualities in general). For flexible response-dependent properties (e.g. moral values), the biconditionals contain no rigidifiers, and do not allow the concepts to latch on to real properties. For these, the aim of vindicating a moderate realism cannot be reached. In the levels terminology, López de Sa's line of

¹⁵ Wright (1992), p. 116.

¹⁶ More precisely, this is the line he says in (1992, p. 116) that he took in the manuscript 'Notes on Basic Equations' that the presentation in (1992) is based on.

¹⁷ López de Sa (2003) and (2006).

thought might be paraphrased as follows:¹⁸ rigidification ties the concept to a low level property (or complex of low level properties) – the one that is responsible for (actual) responses in (actual) C-conditions. This fits nicely with de Sa’s claim that rigidification must be employed across the board – on all of the C-conditions, the subjects and the responses – if employed at all. For non-rigid response-dependence accounts, on the other hand, the lack of rigidifiers mean that the response-dependent concept is not tied to any ‘real’, i.e. low level, property or complex of properties, and hence realism is not vindicated.

It is certainly true that rigidification can be employed in such a way that they mirror intuitions about connections to – or, better, supervenience upon – physical qualities. This much could be true even for colours, as long as the supervenience base is chosen appropriately (i.e. possibly including the laws of nature, and, on some views, the perceptual apparatus of the appropriate subjects). However, rigidification can yield very different results when employed in different ways, and does not entail that the property has to follow a lower level referent across possible worlds.

Wright’s second line of response to the problem about rigidification is a response along these lines. The idea is that rigidification does not have to be done across the board, but only where it is needed in order to respect core intuitions about the subject matter under discussion. For colours, for example, rigidification regarding ‘subjects with actually typical visual apparatus’ is necessary in order to secure that the colours would remain the same if we were to change, say due to an epidemic or a mass trichromate genocide arranged by a jealous dichromate dictator. But other aspects – the laws of nature, maybe – might be allowed to change, and this would sever the connection between responses in C-conditions and the physical properties actually responsible for these. Such local rigidification does not entail low level location.

This seems right, but it is questionable whether it would solve the problem. A first apparent problem is this: if the motivation to rigidify derives from the consideration cited by Wright that rigidified specifications seem the only way to simultaneously respect the substantiality and necessity requirements, then how

¹⁸ Lopez de Sa sounded positive about this paraphrase, but it is not the way he himself presents the matter.

should the non-rigidified parts of the specifications go in order to respect these requirements? If they are made in physical (low level) terms, then this would seem to raise the original problem about necessity, as the low level story might have been different. If they are made in a whatever-it-takes way, this raises a worry about substantiality. This objection might be met by arguing that the substantiality requirement would be met as long as there are any substantial (rigidified) parts in the specifications, since these would make the equations non-trivial even if the rest of the specification is a ‘whatever-it-takes’ formulation.

A second problem with part-way rigidification arises when we view the situation in the light of the levels framework. Response-dependence of subject matter is naturally viewed as a high level location. The effect of rigidification, when done across the board, is to tie the concept to a low level property. Partial rigidification would seem to place the resulting theses between levels in an undesirable way – neither consistently high-level nor consistently low-level. In the context of the levels framework, any intuitions tying the disputed subject matter to lower level properties – exactly those intuitions that rigidification is meant to satisfy – may often be more naturally viewed as intuitions pointing to a low level location, and hence competing with response-dependence theses rather than dictating their shape. Many response-dependent concepts presumably come with such intuitions beside the ones that suggest response-dependence (more on this below). But partial rigidification does not seem the right way to accommodate them; it seems like an unstable compromise. These considerations will have little force for someone who does not endorse the level-based way of thinking about the matter. Such people might rest content with partial rigidification. But for those who do, the strategy won’t help.

A third approach to the rigidification problem is to argue that the necessity in the response-dependent cases is of a more interesting sort than the necessity created by rigidification. Johnston (1998) takes this line and adds a requirement that ‘the canonical biconditionals are not merely superficial necessity produced by ‘rigidifying’ on a relation which is itself contingent’.¹⁹ Wright’s current view on the matter (conveyed in discussion) is also a ‘special kind of necessity’ approach: He

¹⁹ Johnston (1998), p. 10.

holds that the mark of response-dependence is a priori, conceptual necessity – i.e. that the equations are necessary in virtue of the content of the concepts, and not as a result of other factors, such as rigidification. Ralph Wedgwood has pursued a somewhat similar strategy, though it is formulated in terms of properties rather than concepts (as he takes response-dependence to be a property of properties, not concepts). Wedgwood argues that response-dependent properties are properties with response-dependent essences, and that the appropriate biconditionals are *necessarily true in virtue of the essence of the properties*. For response-independent properties, similar biconditionals may well be true, and even necessary, e.g. as a result of rigidification, but not in virtue of the nature of the properties themselves.²⁰

These suggestions are interesting. But they require further work: each of them needs to be backed up by a story of how we can distinguish the interesting kinds of necessity – necessity in virtue of the essences of the properties, or conceptual necessity, or whatever – from the uninteresting necessity that derives from rigidification. In short, there is an issue of how to make the revised necessity conditions operational. It may well be possible to meet this challenge for one or more of the proposals. For example, it might be that the different kinds of necessity give rise to different intuitions about the conceivability of certain scenarios. Or it might be that (other) considerations about how the concepts work can help,²¹ e.g. the considerations behind standard location arguments such as those considered in Ch. 2. But if such criteria can be found, a further question arises: could these criteria (whatever they are) that serve as a sign of the interesting form of necessity also serve directly as a criterion for response-dependence? If so – and it seems likely that this would be the case – presumably we might as well go straight to these considerations and let them provide criteria of response-dependence, rather than taking the detour

²⁰ Wedgwood (1998). This strategy depends heavily on having an account of what essences are and how they are recognised. Accordingly, Wedgwood devotes most of his paper to these issues and argues that essence should be taken as a basic notion, and modal notions are to be explained/defined in terms of essences rather than the other way around. This suggestion is interesting, but I won't explore it further, as the needed investigation into essences would take us too far off topic.

²¹ As Wedgwood suggests; (1998), p. 51.

via essences or conceptual necessity. We shall return to this point later in this chapter.

Our discussion of necessity and rigidification has been less than conclusive. The most promising strategy (at least from my point of view) seems to be the 'different kinds of necessity' strategy. It may well be that we can find symptoms by which to discriminate between the necessity that results from rigidification and the necessity that results from response-dependence. However, if such symptoms can be found, there is a question whether we need to rely on the (new, restricted) necessity condition in order to tell response-dependent from response-independent subject matters, or whether we might as well go straight to these symptoms in order to get an operational distinction.

8. The a priority condition

In the early days of the response-dependence distinction, it was thought that what distinguishes concepts of response-dependent subject matters from response-independent ones is the *a priori* of substantially specified response-dependence biconditionals.²² However, as noted, Pettit argues that even his brand of location-neutral, ethocentric response-dependence can give rise to a priori true biconditionals as a result of the way the reference of the concepts is fixed. If a concept works as described by the ethocentric story, then a property will only qualify as the referent of the concept if it gives rise to the characteristic responses in C-conditions. So even if the concept refers to a substantial, response-independent property, it will still be a priori that things that have this property, and hence fall under the concept, will elicit the appropriate responses in C-conditions. So the appropriate biconditionals will be true a priori. If Pettit is right, the role of the a priority condition needs to be reconsidered. Then the 'standard package' of a priori and substantially specified biconditionals can't be used as an indicator of response-dependence of subject matter, though it does indicate that some sort of response-dependence (of subject matter or of concepts only) is at play.

²² See Johnston (1989) and Wright's account of Johnston's seminar presentation, Wright (1992), p. 109, fn.16.

Johnston has been aware of this complication from the beginning. In (1989), he defines response-dependence in terms of a priori and substantially formulated biconditionals, then adds in a footnote:

At least this holds with one proviso having to do with concepts introduced by reference-fixing descriptions and for which we have an a priori guarantee that there is some natural similarity underlying the relevant sample. Everyday terms for shapes might provide some examples.²³

In (1993), he gets around the problem by defining response-dependence in terms of conceptual identities,²⁴ and in (1998) by adding a necessity condition to the account.²⁵

Wright does not mention the problem, perhaps because his other conditions, the independence and extremal conditions, are not met by concepts that are only response-dependent in Pettit's sense, and so might do the work originally attributed to the a priority condition, or perhaps because he doubts that concepts of the latter kind can really underwrite a priori true, substantially formulated biconditionals (see below).

The first question we should consider in our assessment of the a priority condition is this: Is it true that merely Pettit-style response-dependence can guarantee the a priori truth of substantially specified response-dependence biconditionals? If not, there may be no problem at all. We shall consider two reasons to doubt the claim that Pettit-style response-dependence generates a priori true biconditionals of the usual kind. The first concerns the possibility of reference failure, i.e. situations where no *x* is *F*, but where nevertheless there are things that elicit characteristic responses in *C*-conditions. The second is based on the consideration that for ethocentric concepts with substantial referents, there may be rare cases where something has that

²³ Johnston (1989), p. 146, fn. 8.

²⁴ 'The concept of being red = the concept of the disposition to look red to standard perceivers under standard conditions' etc. (Johnston 1993, p. 104.) This won't be true if the concept refers to something more substantial than a response-disposition.

²⁵ Johnston (1998), p. 9. As response-independent referents might come apart from responses across possible worlds, this condition will not be met if the a priority of the biconditional is a result of reference fixing and not of genuine response-dependence of subject matter.

property, but where the responses would fail to manifest even in the best of conditions (or conversely). It turns out that the first objection can be met, whereas the second calls for adjustment of the equations. So mere Pettit-style response-dependence can't guarantee the a priori truth of the traditional biconditionals. This result, however, does not completely clear the a priority condition of the charges.

8.1 A priori biconditionals and the possibility of reference failure

A first problem for the idea of a priority by reference-fixing arises from the possibility of reference failure. For a concept that is intended to refer to a substantial property via some sort of reference-fixing story, there is always a risk that no property qualifies as referent because nothing fits the 'job description'. It might be that the things that elicit the appropriate responses in C-conditions are very different and have nothing in common apart from eliciting the responses. If the concept fails of reference, nothing will be F, and so the biconditional will fail right to left for any x that elicits the appropriate responses in C-conditions. But whether or not a suitable referent property exists is an empirical matter. In that case, how can it be a priori that the biconditional holds?

We shall examine three possible responses to this challenge. One is to adjust the formulation of the biconditionals in such a way that the right-hand-side will not be true if the concept fails of reference. First, we might reformulate the biconditional as follows:

$$(4) \quad x \text{ is } F \leftrightarrow x \text{ is such as to elicit } R \text{ from } S \text{ in } C.^{26}$$

The idea would be that if the responses would occur, but not through any particular intrinsic properties of x, then that is not enough to make the right hand side true. The 'is such as to' is meant to imply that there is a property of x that is responsible for systematically eliciting R in S in C. This will be true only if 'F' has a referent; to qualify as a referent of 'F' is exactly to be as specified in the right hand side. So the left hand side will never be true when the right hand side isn't. (At least, this would

²⁶ Blackburn (1993) employs biconditionals of exactly this form.

be the case on the reading of the 'is such as to' clause that we've considered. It might also be given a level-flexible reading; for this option, see below.)

This response is problematic for several reasons. One is that if 'F' fails to refer, and if this means that the concept is in bad standing (as is the case for e.g. phlogiston),²⁷ the left hand side will be either false or meaningless, depending on your view about what to say about concepts that fail of reference. If 'F' is meaningless, the left hand side of the biconditional will be likewise. And it is, if not clearly false, then at least controversial whether we should say that a biconditional with a meaningless left hand side and a false right hand side is true.

This problem might be solved by reformulating the biconditionals in such a way as to *mention*, but not *use*, the disputed concept on the left hand side, e.g. as follows:

(5) *x falls under the concept 'F'* \leftrightarrow x is such as to elicit R from S in C

If 'F' fails to refer (and if this means that the concept is in bad standing), then both sides of this biconditional will be false for all choices of F and x. So no instances of the biconditional would be false. We thus get a biconditional where both sides are true of ethocentrically acquired concepts that happen to have substantial referents, and neither side is true for concepts that fail of substantial reference. This much can be known a priori, even though it can't be known a priori which option is relevant for a given concept.

If this is the correct solution to the problem, this would be good news for adherents of response-dependence of subject matter. These theories could hold on to equations of the form

(6) x is F \leftrightarrow x would elicit R from S in C,

and the a priority of such equations might then serve as criterion for response-dependence of subject matter, whereas Pettit-style response-dependence would only underwrite a priori biconditionals of the (5) type. This would be a very neat solution to our problem about making the distinctions operational, and would mirror the fact

²⁷ If the concept is level-flexible, this problem won't arise; for this possibility, see later.

that response-dependence theses of subject matter are indeed theses about the subject matter, whereas Pettit-style response-dependence is a thesis about the concepts only.

There are two problems with this neat solution, though. First, as I shall argue below, if we think about the concepts as level-flexible, even the Pettit-style response-dependence theses underwrite a priori equations in which the concept is used rather than mentioned on the left hand side.

Secondly, and relatedly, one might feel that any response based on saving the biconditionals by having both sides come out false for the problematic cases is disappointing in a certain respect: It threatens to undermine the motivation for using response-dependence biconditionals. The biconditionals were supposed to mirror the ethocentric story and the way concepts are formed by way of responses and the discounting practice that gives rise to the C-conditions. But now we get the result that both sides of the biconditionals are true only for ethocentrically acquired concepts *with substantial referents*, whereas both sides are false for other ethocentric concepts (those that are about response-dependent subject matters) – as false as they are for properties that are totally unrelated to responses in C-conditions. So the biconditional no longer says what is special about the class of ethocentric concepts, and consequently does not capture the sense of response-dependence we were after: the sense that concerns the way the concepts are acquired, and is neutral on the nature of the referents.²⁸

Pettit would presumably resist this line of thought (based as it is on the idea of ethocentric concepts not being low-level rigid). He has suggested (in conversation) that in cases where our use of a concept does not track an independently existing property, the practice with the concept will eventually break down. Where the ethocentric concept formation mechanism works, and gives rise to a practice with the required degree of constancy in responses, it is because we track a substantial property. So cases of reference failure constitute no threat to the biconditionals for ethocentric concepts; they only concern instances that should, and eventually would, be given up anyway. This is a second response to the original challenge.

²⁸ The same objection applies to the strategy based on formulating the left hand side as 'x is *denominably* F' (Pettit 1999, p. 37-38, though this discussion does not directly address the possibility of reference failure).

It is probably true that in the case of most ethocentric concepts, we are on track of something of which an interesting lower level story can be told. In some cases, however, this story is long and complicated, and arguably sometimes too long and complicated for a low level location to make sense (as is probably the case for colours). So it does not seem likely that all the ethocentric concepts are properly construed as referring to substantial low-level properties. More importantly, it is an empirical question whether all ethocentric concepts either refer or would ultimately break down because of lack of constancy etc. Therefore, the hypothesis that they would seems insufficient as a basis for a philosophical theory.

I prefer a third response to the problem, which is based on the idea of level-flexibility. This response probably relies too much on the levels framework to be acceptable to Pettit. Unlike the other responses, however, it accommodates the useful idea that ethocentric response-dependence is compatible with many locations, and not only with reference to low-level, substantial properties. For a level-flexible concept, the correct response to failure of low level reference is not (or not often) a conclusion of reference failure, but the conclusion that a referent should be sought on a higher level. When retreat to higher levels is a fallback option, aiming for a low level referent does not carry with it a risk that the concept may be in bad standing, and thus it does not threaten the a priority of the biconditionals. If a suitable low level common factor can be found, this property can serve as the referent of the concept (provided that this fits with the core intuitions or whatever else goes into making the concept what it is), and it will elicit the appropriate responses because otherwise it would not have qualified as the referent. If no low level referent can be found, the referent can be sought on a higher level – response-dispositions or perhaps the deflationary properties correlated with criterially governed concepts. Either way, the biconditional will be true, and this can be known a priori, though we don't yet know which of the options will make it true.²⁹ This would solve the problem about

²⁹ This story would work even if all ethocentric concepts should turn out to be high-level ones, even though in that case the idea of level flexibility would lose some of its attraction. The same story would be helpful in showing how the a posteriori route to response-dependence of subject matter could give rise to a priori true biconditionals (more on this below).

reconciling a priori biconditionals with the possibility of reference failure, on the condition that *all* ethocentric concepts are (either high-level-rigid or) level-flexible. This seems a fair assumption if they are acquired (or better: evolved) in accordance with the ethocentric story, neutral as it is on the location issue, and requiring no clues to the answer to the location question.³⁰

Note that this response, if correct, would save the a priority of biconditionals like (6) where the disputed concepts are used, and not only those like (5), where they are only mentioned. So if the idea of level flexibility solves the problem, our discussion suggests that biconditionals of this form will *not* serve as a mark of response-dependence of subject matter.

If you are opposed to the idea of level flexibility and think that all ethocentric concepts are either determinately about substantial properties or high-level rigid and about response-dependent subject matters, you might stick to the first strategy discussed. Then the conclusion will be that only response-dependence of subject matter will give rise to a priori biconditionals like (6), in which the disputed concepts are used, while Pettit-style response-dependent concepts aimed at low level reference will only give rise to a priori biconditionals like (5) that mention, but do not use, the disputed concepts.

But the presupposition does not seem reasonable. Level flexibility seems a very natural interpretation of ethocentric concepts, and a prerequisite for all the advantages discussed in Ch. 5. Also, it seems likely that for many concepts, we simply haven't settled whether they are aimed at reference to substantial properties, to less substantial properties, or meant to be criterially governed; we simply leave the question open and aim for the best we can get. So I will stick with the response based on level-flexibility, and will consequently have to do without the neat way of

³⁰ For people who don't believe in levels in the way I have presented them, the idea of level-flexibility may be paraphrased as follows: Some concepts work in such a way that they refer to a natural kind if one is available, or to a more complex physical property if a suitable one is available, or to a response-disposition, or else the concept is criterial. A suggestion along these lines was made by Jussi Haukioja in his talk at the European Conference of Analytic Philosophy, Lisbon, August 2005.

distinguishing between Pettit-style response-dependence and response-dependence of subject matter outlined above.

8.2 Flawed response-dependent concepts and a priority

It is possible even for level-flexible and high-level-rigid concepts, including concepts of supposed response-dependent subject matters, to be in bad standing. But the manner in which such concepts might fail is different from the manner in which (low-level-rigid) referentially governed concepts might fail, i.e. because it turns out that the concept fails to refer to a substantial property in the world. Level-flexible and high-level-rigid response-dependent concepts might fail if it turns out that no consistent set of C-conditions lead to sufficient constancy in the responses for a consistent practice to work. Or, more generally, they can fail because there are no realisers of the role of C-conditions.

In such cases, the situation would seem to be exactly as with concepts allegedly in bad standing because of reference failure: If the concepts in question are used rather than mentioned in the biconditionals, their left hand sides will be false or meaningless.³¹ So doesn't the same challenge arise: how can biconditionals be a priori if it might turn out to be false for some cases, and we don't know which ones are which?

One way to respond to this problem would be the one suggested for Pettit-style response-dependence above: to replace biconditionals like (6) with biconditionals that only mention, but do not use, the disputed concepts. In this way, the biconditionals for the problematic cases will have false left hand sides, as nothing falls under the concepts. And the right hand sides (x would elicit R from S in C, or maybe the 'is such as to' version given a level-flexible reading) will be false as well, since nothing would elicit the right responses in C-conditions that are not well defined.

Both sides of the appropriate biconditionals will be true for all concepts of response-dependent subject matters that are not flawed, and false for those that are

³¹ Provisoed biconditionals might fare slightly better than basic equations; they would have false antecedents and consequents with meaningless left hand sides. But there is still an issue of what to say about such cases.

flawed. There won't be a problem parallel to the worry that the biconditionals will no longer say what is special about ethocentric concepts in general, but only those with substantial referents, as in this case we wouldn't have wanted the flawed concepts to be included in the class of concepts for which we seek an account.

However, it would seem a shame if biconditionals of the (6) type would need to be given up altogether; since they are openly about subject matters rather than concepts, they seem an appropriate expression for response-dependence of subject matter. (And to the extent they would work for ethocentric concepts as well, they would do so because these concepts, too, keep open the possibility of response-dependence of subject matter.)

Fortunately, there is no need to take this route, as a different response is available. A priority does not entail infallibility. In the case of a flawed response-dependent concept, it turns out that a biconditional that we thought to be a priori true is in fact meaningless (if that's how we should think about statements involving flawed concepts). But this does not threaten the a priority of similar biconditionals for concepts that are not flawed. It just means what we already knew: that our beliefs about a priority and a priori truths are fallible.³²

8.3 The problem of Siberian strange-water

If ethocentric concepts aimed at low-level reference are level-flexible, the possibility of low-level reference-failure does not pose a threat to the a priority of the biconditionals – even biconditionals in which the disputed concepts are used. So we are still left with the original problem: a priority of these biconditionals is not a tell-tale sign of response-dependence of subject matter. However, there is a second problem with the a priority of the biconditionals for ethocentric concepts that refer to

³² Why can't the same response be given in the case of Pettit-style response-dependence and reference failure? Because the issue of (low-level) reference failure is a blatantly empirical one, and if the truth of the biconditionals is hostage to such a feature, then they can't be a priori. Also, given my view that ethocentric concepts are level-flexible ones, our biconditionals would simply not capture the right class of concepts if they were such as to be a priori true for ethocentric concepts with low-level referents, but false for ethocentric concepts without low-level referents.

response-independent properties – a problem that calls for modification of the biconditionals at best.

The problem is this: *If a concept refers to a substantial, response-independent property, there can be no a priori guarantee that judgements in (substantially specified) C-conditions will coincide completely with the extension of this property.* Viewing the biconditional as a reference-fixing description ensures that a property has to coincide *roughly* with responses in C-conditions in order to deserve the job as referent of the concept. But if *complete* coincidence is required, very few concepts will presumably come out as referring to substantial properties. Moreover, no matter how perfect the overlap for the cases we know, there is always the possibility that deviant cases appear.

Wright has stated the problem in terms of a helpful example. Suppose that ‘water’ is a basic, ethocentrically acquired concept, and that H₂O is the property that normally brings about the water symptoms in C-conditions. Suppose also that somewhere in Siberia there is a small amount of very strange water – stuff that is H₂O, but does not have the usual effects on our sensory apparatus, i.e. does not exhibit any water symptoms. In such a case, the property of being H₂O presumably still deserves the reference of the concept ‘water’. After all, in the overwhelming majority of cases, H₂O meets the water criteria; it is just these few centilitres in Siberia that create trouble. So we seem to have a case of water that does not appear like water even in the most conducive of conditions. Furthermore, we can’t rule out that this scenario is actual, since we haven’t checked the nature of each and every bit of matter in Siberia and elsewhere. So even if we’re lucky to inhabit a world where all water is well behaved, we can’t know this a priori. So the biconditional can’t hold a priori after all.

Some might object to the example of water, as it is not a paradigmatic example of an ethocentrically acquired concept. It was chosen because the paradigmatic examples, e.g. colours, would not serve our purposes, as these are likely candidates for response-dependence in a stronger sense as well. But the specific example is not crucial. We can give a general recipe for similar counterexamples: Take an ethocentric concept with a response-independent property as referent, and suppose that the property almost always gives rise to the characteristic responses in C-

conditions, except for a few examples of things that share what we take to be the essence of the property (belong to the same natural kind or whatever), yet consistently fail to display the symptoms (responses), even in the most conducive conditions. This scenario could be real, in which case the biconditional will fail left to right. Even if it isn't real, we can't know this a priori, and so can't know the biconditional to be true a priori. Any counterexample of this sort will do.³³

8.4 The parallel problem of fool's water

A parallel problem can be raised for the right to left reading of the biconditionals. There might be a substance – fool's water – that would consistently meet the water criteria even in the best of conditions, but which is so different from H₂O that we would like to say that it is not water (e.g. because it has a very different chemical structure and some uses that are very different from those of water). In such cases, the right-hand-side would be true, but the left-hand-side false.

The two types of counterexamples are both examples of one overall problem: the lack of a guarantee for a complete fit between a reference-fixing response-pattern and a response-independent referent. In the following, we shall consider a range of possible responses to the problem. We shall focus on the case of Siberian strange-water, but the results carry over to fool's water, and to cases of strange-F and fool's F in general.

8.5 Strange-water-proofing strategies

A first cluster of responses to the problem of strange-water would be based on the thought that *strange-water is not properly viewed as a case of water that is not recognised by responses in C-conditions*.

One version of this strategy would be to claim low level reference failure if there are any deviant cases at all, however peripheral, and retreat to a higher level location for such concepts. Then the strange-water would have no claim to be water,

³³ Of course, if you think that only concepts of response-dependent properties can be acquired ethocentrically, there will be no such counterexamples; Siberian strange-water is a problem only for Pettit-style response-dependent concepts with response-independent referents.

even if it is H₂O. This, however, seems a very unlikely account of what we would do when faced with a sample of Siberian strange-water. There may be some concepts for which we would require total fit between a property and the response-pattern in order for the property to qualify as referent for the concept. But this is not likely to be the general situation; most concepts where low level reference is in the picture – even those that seem good candidates for ethocentrically acquired concepts – simply seem to allow some scope for strange cases, as long as they are not too widespread or troublesome.

A better version of the ‘strange-water is not water’ strategy is based on the idea that if there is strange-water that doesn’t display water symptoms, there must be a good lower level story of how it is different from other water, and hence a good reason for doing either of two things:

Either we could say that the strange cases are sufficiently different from normal ones in their low-level features to not count as falling under the concept after all. The conclusion would be that, surprisingly, ‘water’ doesn’t refer to H₂O, but to the (low level) property that unites the familiar, well behaved water cases, and which does not include Siberian Strange-water (which is really not water at all). This, like the first response, would entail that our assumptions about the referents of basic concepts would be more fallible than we would ordinarily have thought; for all we know, we could be wrong about the referent of ‘water’ (and would be wrong if the world contained Siberian Strange-water), even assuming that everything we now know about chemistry is correct. This seems odd. (But of course this intuition might be there because our concept of water is not an ethocentric, level flexible concept, but has a strong element of level rigidity now that we know about the low-level property correlated with it.)

Or, alternatively, we could say that in the light of an interesting new discovery, we have to revise the concept ‘water’: we had the ethocentrically acquired concept, and though pointing us to a natural kind, H₂O, it is not completely accurate. So on the basis of this concept, we should introduce a new, non-basic and low-level rigid concept that includes all instances of H₂O, strange and normal, and which is defined by a scientific story and not by the ethocentric story. Beside the new concept, we might keep the old water concept as a higher level one, e.g. a criterially governed

concept referring to everything that displays the water symptoms. This strategy, too, presupposes willingness to claim low-level reference failure for a concept with an ‘almost always’ low level correlate because of the discovery of a few strange cases. This seems counterintuitive.

All the alternative versions of the ‘strange-water-is-not-water’-strategy seem out of step with what we would be likely to think and do about strange cases. So we should look for a better alternative,.

A second, different approach to the problem would be to *adjust the biconditional to make provision for a few strange cases*. In order for a property to deserve the reference of a concept, we need not require complete correlation with responses in C-conditions, or whatever the reference-fixing description is about. All that is needed is sufficient constancy in responses to get a concept off the ground, and enough correlation between the candidate property and responses in C-conditions for the concept to qualify as a referent of the concept. It seems intuitively reasonable to say that water refers to H₂O even if there happens to be a couple of centilitres of strangewater somewhere in Siberia. So while we can’t know a priori that the biconditional will be true for each and every case, what we can know a priori is that it will be true *as a rough and ready generalisation*. We might try to capture this by reformulating the biconditional along the following lines:

(4) For most cases: x is F \leftrightarrow x would elicit R from S in C

The ‘for most cases’ clause must be in a proviso; if it was incorporated in the right hand side of the biconditional, it would not handle cases of fool’s F, and if it was part of the left hand side, it would not handle cases of strange-F.

The big question about such a clause is what sort of cases the quantifier ‘most’ ranges over. It should not be all possible choices of x ; there might be worlds where very different things elicit R from S in C, but this should not prevent an actual common factor from being the referent, unless there are independent reasons that this could not be the case (e.g. strong, durable transparency intuitions). It should not even be all actual cases of F; it might be that 90% of all actual water is strange-water, though the water on earth is H₂O. It might even be that 90% of the earth’s water is strange-water, but that it is a little heavier than ‘normal’ water, and hence is found

only deep in the ground and oceans, and is rarely encountered by humans. The relevant cases must be the ones we come across – those that have helped form the concept, and those that continue to do so in our practice with the concept: most cases of evaluation of an x by an S in C . This makes the equation very different from the usual response-dependence equations that are not restricted to cases we come across.

Another problematic issue about the proposal is the proportion and distribution of cases needed for a property to deserve the role as referent for a concept. Presumably it varies a lot from concept to concept how much overlap is required, and presumably there are core cases that are intuitively very important, and peripheral cases that are less important. So a general recipe is unlikely to be forthcoming, except maybe as a ‘role-specification’ along the lines of ‘enough and important enough cases for the property to be acceptable as referent’.

It may be possible to find a modification of the biconditional that will do the job, but as this suggestion will play a very limited role in the following, I shall not go further into the complications surrounding it. We should note, however, that if this line is taken, the ‘in most cases’ clause will mark a difference between response-dependence of subject matter and mere ethocentric response-dependence. The a priority of biconditionals without this clause might then serve as a test for response-dependence of subject matter.

It might be argued that a situation parallel to that of Siberian strange-water could arise for response-dependence theses of subject matter as well. Even for dispositional concepts, there can be cases where the expected response fails to manifest for no good reason, even though the conditions are as good as can be. For example, one in a million times a fragile glass does not break when dropped onto a hard surface from a height that would normally be sufficient to break the glass (say, has broken 999 others from the same production line).³⁴ In such cases, we don’t conclude that the glass isn’t fragile after all (you wouldn’t push your luck by dropping it once more), or that something was wrong with the conditions; we take it

³⁴ This is different from the problem cases of masking and mimicking to be discussed in Ch. 8; in those cases, the unexpected events are results of interference of some sort, and don’t happen for no good reason.

as just plain luck. This suggests that even with purely dispositional concepts, the ‘in most cases’ or something similar might be needed, explicitly or implicitly. (Our example is not a response-dispositional concept, but this shouldn’t make a difference.) So the need for a proviso of this kind may not be restricted to response-dependent concepts with response-independent referents, and so the original biconditionals might not serve as markers of response-dependence of subject matter after all.

If the problem of Siberian strange-water is seen as a challenge from adherents of response-dependence of subject matter against Pettit’s view (‘leave our biconditionals alone!’), then this could be taken as a third possible reply to the challenge: Whatever the problem is for Pettit-style response-dependence, it would arise for other response-dependence views as well.

This response does not seem convincing, as there are important disanalogies between the problem cases for the two kinds of view. One crucial difference is that in the problem cases for Pettit-response-dependence, e.g. Siberian strange-water, *the samples themselves* are such that they would never elicit the relevant responses, whereas the fragile glass that doesn’t break on one occasion would be likely to do so if given another chance.³⁵ Another is that with Pettit-response-dependence, there should intuitively be room for deviant cases on a larger scale than with response-dependence of subject matter (since there is a property with an independently given extension to guide the extension of the concept independently of responses in C-conditions, whereas in the case of response-dependence of subject matter, there is nothing other than responses in C-conditions to appeal to). Finally, there are differences in the way the C-conditions work in the two cases, and corresponding differences in the way we would treat deviant cases; we shall return to this issue in Ch. 7.

A fourth possible response to the problem of strange- and fool’s water would be to concede the point and give up the a priori biconditionals. For a Pettit-style

³⁵ This is also why (some of) the defence strategies against conditional fallacy problems to be considered in Ch. 8 would not solve the problem with deviant cases for Pettit-response-dependence, but would do so for the other forms.

view, this would not be a devastating loss. It is questionable what work the biconditionals do for such a view anyway; the real work is done by the ethocentric story of concept acquisition, and the biconditionals are employed because they are thought to mirror that story. But the biconditionals are not particularly helpful in capturing the intuitive content of a Pettit-style response-dependence thesis if they don't distinguish between this and other versions of response-dependence (except possibly by the need for an 'in most cases' clause or similar). Furthermore, the biconditionals are not needed for making the distinction between response-dependent and –independent concepts operational, as Pettit's brand of response-dependence is supposed to apply globally. They would presumably not be helpful for distinguishing basic from non-basic concepts either; for this purpose, considerations about the way the concepts work, how they are acquired, and how they might have developed seem more appropriate. Pettit could thus give up the biconditionals without losing anything of importance.

We can conclude that the objections to the a priority of the biconditionals on Pettit's version of response-dependence are not devastating to the view. First, Pettit and his followers could dispose of the biconditionals and let the ethocentric story do the work. Secondly, it is possible that the problems could be solved by adjustments to the biconditional that would make them consistent with low-level referents and the resulting possibility of 'strange cases' – adjustments which might even be needed for other versions of response-dependence as well.

Before we can draw a conclusion with respect to our main question about the a priority condition, we shall consider a final relevant issue: how the a posteriori route to response-dependence considered in the beginning of this chapter combines with the a priority condition.

8.6 A priori biconditionals and the a posteriori route to response-dependence

If a conclusion of response-dependence of subject matter is reached by a discovery of low level reference failure of level-flexible, ethocentric concepts, can the biconditionals still be a priori? It is hard to see how they could, except for the same reasons – and to the same extent – as the biconditionals can be a priori for ethocentric (Pettit-style) response-dependence, i.e. as follows:

- 1) if the concept refers to a response-independent property, that property will be picked out by responses in C-conditions, and so will make the biconditional true, though possibly in a version adapted to accommodate strange- and fool's cases
- 2) If there is no appropriate low level referent, the reference should be located on a higher level with response-dispositions or with the deflationary properties correlated with Euthyphronic concepts. This, too, would make the adapted biconditional true, and would also make true one of the original biconditionals

We could describe the situation by saying that the (sentence stating the) adjusted biconditional can be taken to state two different propositions: One about reference fixing, corresponding to 1), and one about response-dependence of subject matter, corresponding to 2). Either of them would make the biconditional true. If we let the choice between these options depend on the answer to the location question, we can know a priori that the biconditional will turn out true. Of course, the biconditional must be equipped with any patches needed to accommodate the possibility of strange-F and fool's F, in case the relevant alternative is 1). In this way, it might be argued that even response-dependence by the a posteriori route might give rise to a priori true biconditionals, though only of the kind that are appropriate for a reference-fixing story.³⁶

There is also something else that we can know a priori: *If* a concept is a concept of a response-dependent subject matter, even by the a posteriori route, then a biconditional of the usual kind holds for it. And since the concept fails of low level reference, there will be no analogues of Siberian Strange-water or fool's water.

8.7 Summing up the discussion of the a priority condition

Our discussion, though somewhat inconclusive, has yielded a part-way vindication of a priori biconditionals as a test separating different sorts of response-dependence

³⁶ This point may raise objections. For example, if the adjustments needed for Pettit-style response-dependence are not consistent with response-dependence of subject matter, no biconditionals will be a priori for level-flexible concepts. However, as I am about to argue that the a priority condition might be replaced with other considerations, not much will hinge on this issue. So I shall leave the matter here.

from the corresponding forms of response-independence. The problem of reference failure can be solved in two ways, depending on whether you think of ethocentric concepts as level-flexible. If you do, then the level-flexibility of the concepts enable them to make response-dependence biconditionals – even those where the disputed concept is used on the left hand side – a priori true. If you don't believe in level-flexibility, the conclusion will be that biconditionals in which the disputed concepts are *used* on the left-hand-side will only be a priori for response-dependent subject matters, while biconditionals in which the disputed concepts are only *mentioned* will be a priori even for Pettit-style response-dependence. As I think ethocentric concepts are level-flexible, I think the former alternative is the way to go.

Strange- and fool's cases pose another problem for the compatibility of a priori biconditionals with Pettit-style response-dependence. If the correct response to this problem is one that allows even Pettit-style response-dependence to be combined with (adjusted) a priori biconditionals, these will serve as a criterion that some sort of response-dependence is in play. Biconditionals without the adjustments (though with any modifications necessary for dealing with conditional fallacy problems; see Ch. 8) might then serve as a criterion that a 'stronger' form response-dependence is in play.

8.8 Beyond the a priority condition

A natural next question is this: How do we determine what biconditionals are a priori? My main reason for dissatisfaction with the a priority condition has to do with the answer to this question.

An unpromising answer would be that language users should know the biconditional to obtain just in virtue of their competence with the concept. Such a suggestion doesn't make much sense when we are talking about the man in the street, who has better things to do than to concern himself with biconditionals and standard conditions. So claims about a priori biconditionals have to be motivated in another way. The obvious place to look would be intuitions about the concepts, and about what we would say about apparent deviant cases. However, intuitions about such matters are fairly rough and imprecise. Thus, we might not be able to discriminate between biconditionals that mostly hold and those that always hold, or between those

that hold as a result of a priori response-dependence and those that we now know to hold because it has turned out that the concept does not refer to a low level property.

There are ways to get intuitions into clearer focus, however. We have talked about some of them: the cases of strange-water and fool's water are exactly such devices. Such cases may help us to clear away the confusion and distinguish the cases properly.

So much the better for the a priority condition? No, so much the worse. For if intuitions about such cases are what do the work, we might as well go straight to these cases in our attempts to determine which concepts are response-dependent in which senses, rather than taking a detour via an a priority condition.

Intuitions about strange- and fool's water take us quite some way towards this goal. So do intuitions about the possibility of reference failure. A recipe for detection of response-dependence by the means we have considered in the discussion of a priority might look something like this:

If the extension of a concept coincides with responses in C-conditions, there is a good question whether the concept is response-dependent. To determine whether, and in which sense, it is response-dependent, we should consider (among others) the following questions:

- Could this concept turn out to be in bad standing because it does not refer to a substantial property?

If yes, this suggests that the concept is not response-dependent, but low-level rigid. If no, the concept is response-dependent in some sense or other. In what sense? This depends on the answer to the next question:

- Is it conceptually possible that there could be strange-F or fool's F?

If not, this is a sign of response-dependence of subject matter (either by the a priori route, or possibly by the posteriori route, provided that we already know that the concept has no low level referent). It is not a conclusive proof, however, since the verdict could be due merely to lack of imagination on our part. This is a problem, but not one that is specific to this way of determining response-dependence status; it would have been there for the a priority criterion as well.

If strange-F or fool's F is possible (but where low level reference failure would still not have entailed that the concept was in bad standing), the concept is level-

flexible. It might be known to have a low level referent, or its location might be unsettled. The situation is a bit more complicated if the concept is known *not* to have a low level referent, i.e. to be response-dependent by the a posteriori route. In that case, we might have forgotten that we once thought it might have a low-level referent, in which case intuitions will be the same as for high-level-rigid response-dependent concepts. Or we might still think as we did when the location question was unsettled, in which case strange-F and fool's F will still impress us as possible.

Because both possibilities are open for the a posteriori route to response-dependence of subject matter, intuitions about strange-F and fool's F will not completely settle which brand of response-dependence we are dealing with. But it takes us some of the way, and presumably further than the a priori criterion. And luckily, there are other ways of settling this question.

9. Location arguments as criteria of response-dependence

If we think of response-dependence theses of subject matter as theses of high-level location with responses in C-conditions, another natural place to look for clues to response-dependence status suggests itself: the standard argumentative strategies used in location disputes. The types of location arguments considered in Ch. 2 – causal arguments, transparency arguments, and unity arguments – are all relevant. In fact, many of them are already found in more or less explicit form in discussions about response-dependence.

9.1 Causal arguments

Johnston's missing explanation argument is the clearest example of this kind: a causal argument designed specifically for the purpose of distinguishing response-dependent domains from response-independent ones. As it stands, Johnston's argument applies to all response-dependence theses, including Pettit-style ones (at least provided a suitable strange-water-repellent can be found): it applies to any view that employs response-dependence biconditionals with 'a priori equivalencies' between being F and seeming F or similar. The invited conclusion is that such claims are inconsistent with empirical explanations of the form 'x seems F because it is F'. So if the argument was successful, the appropriateness of such explanations for a

domain would be a sign of response-*independence* in all senses, even Pettit's (though defeasibly, as a revisionary response-dependence thesis would still be a possibility). This conclusion would be surprising, as Pettit-response-dependence should not have implications about the sort of subject matter under discussion, and hence we should expect it to be neutral on matters like causal connections. However, as indicated in Ch. 4, both explicit formulations of the missing explanation argument are flawed as they stand, so we need not worry about this strange result. Nevertheless, there is an important point in the vicinity: intuitions about causal efficacy might be expected to be a good guide to response-dependence of subject matter. For response-dependent subject matters, the causal work in bringing about the responses is presumably done by low level properties – the bases, if we are thinking in dispositional terms – rather than by the disputed properties themselves. So strong and durable intuitions about causal efficacy, particularly regarding responses in C-conditions, will be a sign of response-independence of subject matter.

Further distinctions can be made regarding causal intuitions. For example, it seems characteristic of good candidates for response-dependent subject matters like colours that the things they cause are either responses in subjects, or caused via responses in subjects. The colour of a flower might attract insects. The colour of a malfunctioning traffic light might cause an accident via its effects on unwary drivers. But it is hard to find an example where the causation done by colours is not done to, or via, perceivers. By contrast, shapes can cause things without the mediation of minds; the reason that a round peg doesn't fit into a square hole of similar area has nothing to do with observers. Such differences form the basis for Wright's wide cosmological role constraint. This is presented as another test of realism besides the Euthyphro contrast.³⁷ But there are presumably close relations between the two. Wide cosmological role naturally goes with response-independent subject matters, while response-dependent subject matters have narrow cosmological role.

Finally, there may be interesting distinctions to draw between different kinds of response-dependent concepts according to the relation between the relevant response-dispositions and their causal bases. For some dispositions, bases matter a

³⁷ Wright (1992), p. 196-99, which inspired these remarks.

lot, and we expect the manifestation of the disposition to be caused in a particular way by a particular intrinsic property in every instance. For other dispositions, all we seem to care about is that the manifestations would occur, whereas the story about how they come about is irrelevant. Differences like these, if described in more systematic ways, might provide the grounds for further distinctions between different sorts of response-dependence of subject matter. However, this is one of the issues that we shall have to put aside for later investigation.

9.2 Transparency arguments

Transparency considerations also have an important place in the discussion about response-dependence. Strong and durable intuitions about transparency – that if something is F, this will be apparent to a subject, and conversely, at least in favourable conditions – are presumably a primary criterion for response-dependence by the a priori route. By ‘durable’, I mean intuitions that don’t go away once we realise that there are different levels involved, and that we can’t have both transparency and e.g. causal efficacy. A paradigm example would be the transparency intuitions about pain that drive Kripke’s argument against identity theories. Such intuitions may be translated into exactly the kind of a priori biconditionals that are thought to be characteristic of response-dependence.

Transparency intuitions can presumably be respected to a certain degree for level-flexible response-dependent concepts (whatever the correct location turns out to be), for reasons familiar from our discussion of a priority: No property could qualify as the referent of a response-dependent concept unless it would elicit the right responses in C-conditions. For level-flexible concepts, however, a few weird cases should be acceptable; cases of strange-F and fool’s F should be conceivable.

(If they are not, this could be because we simply haven’t considered the matter carefully, and confuse the fact that responses in C-conditions normally get it right with the principle that they always get it right. Or it could be because the concept works in such a way that we would be prepared to claim reference failure if there were any weird cases at all, and should only accept a low level property as referent if the fit with responses in C-conditions was perfect. (This is an option in principle, but it is questionable whether there are such concepts; what could be the motivation for

so strict standards except considerations that would point to a priori response-dependence if they were thought through?))

If a level-flexible concept is recognised as a posteriori to have a response-dependent subject matter, then this knowledge should make for stronger transparency intuitions.

But where transparency intuitions are most important is as a criterion for a priori response-dependence of subject matter. If transparency intuitions apply across worlds in an appropriate way, thus making the physical story – and in particular the question whether an appropriate low level common factor is correlated with the concept in the actual world – irrelevant to the extension of the concept, this is a sign of a priori response-dependence. Questions like the following might help tease out relevant intuitions: 'if the underlying properties were different, could x be F just the same?'; 'with no knowledge of lower levels, could I know the nature of the property?' etc. Positive answers would suggest response-dependence by the a priori route. Again, these are not sure-fire criteria, as the intuitions might result from confusion as well as response-dependence. But it is presumably one of the best indicators we can get.

A further, and related, type of relevant considerations will be what is *the highest court of appeal* regarding the 'property' in question. If responses are intuitively the highest court of appeal, this suggests that the subject matter is response-dependent. If it is conceivable that another criterion – e.g. some sort of scientific investigation – could become the primary detection method for the property in question, and the ultimate test in hard cases, then this suggests a response-independent subject matter, or at least that the concept is such as to be apt for low level reference, and so that a conclusion of response-dependence of subject matter, if appropriate, will be by the a posteriori route. (This criterion suggests that colours are response-dependent by the a priori route; while investigations of reflectance spectra and surface properties can tell us interesting things about the colours that we didn't know before, it seems wrong to say that such detection methods could be the ultimate court of appeal for colour judgements, and trump the deliverances of colour experiences in C-conditions.)

Considerations about the highest court of appeal may belong on a different stage in the investigation, though: it may belong as a part of the intuitive ideas at the core of the response-dependence distinction, rather than as a criterion relevant only for operationality purposes.

9.3 Multiple realisation arguments

Multiple realisation arguments can also serve as arguments for response-dependence of subject matter. Depending on the version, they can support both the a priori and the a posteriori variety. If the point of the argument is that there *could* be multiple realisers, in this world or others, and therefore identification with an actual realiser is out of the question, then this counts in favour of a priori response-dependence of subject matter.³⁸ If the crucial point is that there has *actually* turned out to be multiple realisers rather than an expected low level common factor, it will be a sign of the a posteriori variety (unless, of course, this is combined with the claim that the possibility of this scenario alone rules out low level location, even if we might have missed this until, to our surprise, it turned out that lack of unity on the lower level is compatible with the unity found on a high level; in that case, the response-dependence theses should have been reachable a priori, and this was missed only due to failure to think things through).

9.4 Further problems with operationality

All these considerations are somewhat preliminary. This is because the task of developing them further is one that is best done domain by domain, rather than by way of a general recipe. The various argument types will have different shapes and weight for different domains. Also, there will be different additional constraints to take into account for each domains. For moral values, for example, considerations about normativity and about the connection to motivation will be important as well as considerations like those stated here.³⁹ This is one of the reasons that I do not take

³⁸ Of course, an alternative conclusion would be that the concept refers to different things in different worlds, but this, too, might be out of step with intuitions.

³⁹ Zangwill (2003) argues that response-dependence theses of morality cannot meet these constraints.

on the task of developing a complete recipe for determination of response-dependence status. What I have done is to highlight some features that are relevant in several domains, and which should give an idea about the types of features to look for when considering the question of response-dependence in each domain.

A further complication regarding the idea of finding a general, fool-proof test of response-dependence is the often incomplete, and often contradictory, picture painted by the intuitions and other location-relevant features in the relevant domains. First, many concepts seem to equivocate between levels in the way described in Ch. 2: there are intuitions speaking for (and against) each of various incompatible locations. The intractability of many of the location disputes in which response-dependence theses are live options suggests, in my opinion, that many of the concepts under discussion equivocate between properties on different levels. If this is the situation, no single type of intuition or argument should be taken as conclusive evidence for or against a given location. The location question, and with it the status regarding response-dependence, should be settled by looking at the overall picture, including all location-relevant intuitions and the weighting between them. This is why e.g. the missing explanation argument is misleading if considered as a knock-down argument that settles response-dependence status once and for all. In order to be convincing, it must be part of a larger picture where other, competing intuitions about the subject matter are taken into consideration as well.

I think most concepts that are subject to location disputes spread over more levels than one. Though it makes a lot of sense to distinguish different principled ways a concept might function, e.g. natural kind concepts and Euthyphronic concepts, I also think that most concepts, as they stand, are really mixtures of components of different kinds. This presumably means that 1) there is often no one simple truth about the location of a concept for our tests to reveal, and 2) intuitions will be contradictory and not completely reliable indicators when we apply a test to find out whether a concept is of a given kind. There may be concepts that would pass tests for being referentially governed as well as failing the order of determination test and so coming out as Euthyphronic, because they contain components pointing to different locations.

An objection to this line of thought may be this: If, on my conception of things, most or all familiar concepts come out as confusions, what speaks for the account? Shouldn't we conclude either of two things: Either that the distinctions I make don't cut the cake in the right way – that, as most concepts overlap several levels, how can the principled kinds of concepts I distinguish be the ones that there are (or that it is most useful to distinguish)? Or, alternatively, if my picture is right and most concepts contain level-equivocations, shouldn't we conclude that the concepts should be purified of the equivocations, either by discarding all but one single-level component, or by distinguishing several components that, from now on, we should not mix up? In any case, the thought would be, the confusion can't go on; our current concepts, if they equivocate, are fit for the bin. We thus get the following dilemma: If my division into different levels and different kinds of concepts is right, how come this shouldn't lead to large scale revisionism? If, on the other hand, our concepts are all right as they stand, doesn't that mean that my division must fail, and that we should look for an account that will accommodate all components of the concepts?

My reply to this challenge is that it is natural that the concepts equivocate, given that higher level features are our means of tracking lower level features, even before we have some other sort of access to their nature. We need not, and should not, revise on a larger scale; level-equivocal concepts are well suited for everyday purposes. However, doing philosophy is another matter. For this purpose, it makes sense to try to keep the levels apart, and to get clear about the different ways a concept might function, even if, as usual, real life cases are more messy than the theory suggests.

A further reason why it is hard to set up sure-fire criteria for the various kinds of response-dependence and –independence in terms of intuitions about reference failure, strange-F and fool's F, causal powers, transparency, unity, etc. is semantic indecision. Concepts and the accompanying intuitions may simply be undecided between locations, not because of contradictory intuitions, but because we simply haven't settled what kind of properties the concept is fit to refer to. For many

concepts, it may be that no very firm views or expectations about the concept's location are implicit or explicit in the practice.⁴⁰

This is in keeping with a central point about the ethocentric story of concept evolution: concepts can be in use and in good standing before it is known, or decided, on which level their referents are located. Indeed, it might be taken to be one more advantage of that story that it can accommodate the widespread phenomenon of semantic indecision. The account of response-dependence of subject matter is also strengthened by the fact that we have a (ethocentric) story to tell about how the indecision, and the difficulties with making the distinction operational that it gives rise to, could be there.

There are other theories that would accommodate many of the same types of intuitions as response-dependence accounts. For example, some versions of functionalism seem to do a good job of explaining exactly the intuitions that could also serve as a sign of response-dependence. Role-level functionalism, 'whatever-realiser-functionalism', and two-dimensionalist accounts will all be able to accommodate many of the same intuitions as response-dependence theories. For all these views, it will often (always?) be possible to find intuitions that distinguish the theories and help decide between them if we look at all the relevant intuitions and arguments. This is another reason that the overall picture matters more than particular arguments; sometimes, in order to determine whether a response-dependence thesis is preferable over the alternatives, we must take into consideration a whole range of arguments, intuitions, philosophical presuppositions (e.g. the one about the basic level that seems to be doing a lot in tempting people away from high level location views), and scientific data where appropriate. But again, the best place to consider the particular moves is in particular domains, so I won't try to give a general recipe here.

10. Conclusion

The levels framework offers a useful perspective on response-dependence. First, it allows us to distinguish clearly between Pettit-style response-dependence, which is

⁴⁰ I am indebted to Crispin Wright for a very explicit formulation of this point.

neutral on the location issue and concerns concepts only, and response-dependence of subject matter, which amounts to high-level location of the disputed phenomenon. Secondly, it provides an ‘umbrella formulation’ of response-dependence of subject matter that is consistent with the various versions in the literature, e.g. Johnston’s, Wright’s, Wedgewood’s, and DeClercq’s, as different as they are. Response-dependence theses of subject matter are theses of high-level location – location on the level of properties whose extensions depend on responses in C-conditions.

Next, the levels framework has enabled us to distinguish two routes to response-dependence of subject matter: An a priori route appropriate for high-level-rigid concepts, and an a posteriori route which is relevant when level-flexible concepts turn out to lack suitable low-level referents. In total, we get three categories of response-dependent concepts: Level-flexible, ethocentrically response-dependent concepts with low-level referents, level-flexible concepts with high-level referents, and level-rigid concepts with high-level referents.

Two seemingly important points of disagreement between Wright and Johnston concern the question whether the response-dependence distinction should be formulated in dispositional terms or in terms of subjunctive conditionals, and the question whether response-dependent concepts are best understood as criterially governed concepts or referentially governed concepts with response-dispositions as referents. Against the backdrop of the location dispute about dispositions, and assuming (plausibly) that a high-level location of dispositions is the most suitable choice in the context of response-dependence theses, these disagreements turn out to be less important than they might have seemed. In particular, if dispositions are given a traditional conditional analysis, dispositional and subjunctive formulations are simply equivalent. The disagreement regarding referentiality and criteriality can be explained away as a result of approaching the same subject matter – response-dispositions – from two different angles; a first person perspective on response-dispositions invites a (Wright-style) account in terms of criterially governed concepts, while from a (Johnstonian) third person perspective, response-dispositions look like substantial referents.

In the second half of the chapter, we have considered the question of operationality: Given a subject matter about which we are in doubt, how can we determine whether it is response-dependent?

The traditional answer is ‘A subject matter is response-dependent if it underwrites true biconditionals of the familiar kind which fulfil a set of conditions, including a priority and sometimes necessity’. However, this answer turns out to be unsatisfactory. A first problem is that the a priority and necessity conditions both appear to be satisfiable in two different ways. The necessity of suitable biconditionals can be due to response-dependence of subject matter, but it can also be a by-product of rigidification. The most promising response to this challenge is probably to argue that the relevant necessity can be distinguished from necessity that arises for irrelevant reasons. However, any symptoms that might help with this task could presumably be taken as symptoms of response-dependence as well, thus avoiding the detour via necessity.

For the a priority condition, similar problems arise; Pettit argues that the biconditionals can be a priori as a result of the way the reference of disputed concepts is fixed, and not only because they are about response-dependent subject matters. We have considered two ways to challenge this claim. One is based on the possibility of reference failure. This challenge can be met by arguing that the concepts in question are level-flexible, and thus that the alternative to low-level reference is high-level location rather than reference failure.

The second challenge is based on the thought that if a concept refers to a substantial, response-independent property, there can be no guarantee of complete correlation between the property’s extension and responses in C-conditions; strange- and fool’s cases should be a possibility. It may be possible to meet this challenge by adjusting the biconditionals, e.g. by inserting an ‘in most cases’ proviso, but this suggestion is far from unproblematic. In any case, strange- and fool’s cases are as efficient for deciding response-dependence status as they are for deciding a priority, which means that the detour via the a priority condition becomes unnecessary.

My conclusion is the heretical one that biconditionals and conditions are not the best way to make the response-dependence distinction operational. Instead, we should go straight to intuitions about e.g. strange- and fool’s cases. Given that

response-dependence theses (of subject matter) are high-level locations, the standard argument types in location disputes – causal arguments, transparency arguments, and multiple realisation arguments – can also be employed in deciding response-dependence status. They are already common currency in the debate, as illustrated by the missing explanation argument. However, a general recipe for determining response-dependence status is probably not forthcoming, since the relevant intuitions and arguments can be expected to vary a lot from domain to domain. For each domain, it is important to take the overall picture into consideration, since there may well be conflicting intuitions at play.

I do not suggest that the biconditionals should be given up completely. They can still be useful for setting up the discussion, and particularly for getting an overview of the many points at which there is scope for variation. But for operationality purposes – and indeed for capturing the exact distinctions we are after – a priori and/or necessary biconditionals are not the best way to go.

Ch. 7: Favourable conditions

In this chapter, we shall consider an apparent problem for response-dependence theses regarding the specification of the C-conditions. For response-dependence theses of every description, it is traditionally required that the C-conditions must be specifiable a priori, and in a substantial way that does not trivialise the equations they are part of, but gives them concrete content. These two requirements are prima facie conflicting. I explore a functionalist strategy proposed by Pettit in response to this problem. I discuss some objections to the functionalist approach, and some alternative ways of specifying the role of C-conditions.

While the traditional apparatus of biconditionals and the a priority and substantiality conditions is used in setting up the problems, the conclusions reinforce those from Ch. 6: the conditions traditionally imposed on response-dependence accounts need modification, and are not the best way of making the distinctions operational. I argue that the a priority and substantiality conditions must be understood in a more relaxed way than many do, but that this raises no serious problem for response-dependence theses, as the requirements can all be met to an extent that satisfies the constraints that originally motivated them. The material in this chapter is less developed than that in earlier chapters; it is more of a programme for further investigation than a conclusive argument for a settled view.

1. Substantiality and a priority: the prima facie conflict

As we have seen, response-dependence theses come in very different forms and with very different requirements for a concept to count as response-dependent. But nearly¹ all versions agree on a ‘minimal package’: an *a priori true* and *substantially specified* biconditional or provisoed biconditional. (We have already covered this ground, but I shall briefly sum up the relevant points.)

The substantiality condition says that the C-conditions and the other placeholders in the equations must not be specified in a ‘whatever-it-takes-way’ that makes the equations trivially true, as this would make the equations a priori true even

¹ Except Miscevic (1998).

for response-independent concepts. They must be specified in a way that gives the equations concrete content.

The a priority condition says that a suitable biconditional or provisoed biconditional must be a priori true for the concepts in question. To fix ideas, let's use basic equations for the purposes of this chapter:

- (1) x is F (/falls under the concept 'F') \leftrightarrow x would elicit response R from subjects S in conditions C

The a priority condition is motivated in different ways for different versions of response-dependence. In Wright's version, the motivation is straightforward: If the extension of the disputed concept is determined by responses in C-conditions, then in C-conditions, facts and judgements cannot come apart, and this can be recognised a priori. (I have argued in Ch. 6 that the a priority condition is not the best way to distinguish response-dependent subject matters from response-independent ones, but for all that has been said so far, it should still hold for response-dependence of subject matter.)

For Pettit-style response-dependence, the a priority is supposed to be a result of the way the reference of the concepts is fixed; a property qualifies as the referent of 'F' only if it would elicit the appropriate responses in C-conditions. As we have seen, this will not make the biconditionals a priori for concepts with low-level referents, as provision must be made for a limited proportion of strange- and fool's cases. But they will still hold as rough and ready generalisations, and this much can be known a priori. (And perhaps biconditionals with a 'mostly' clause or similar will be a priori true.) This is enough for the challenge we shall consider to get a grip.

The challenge is this: If the specification of the C-conditions is given concrete content, how can the equations containing them be known to hold a priori?

(Similar questions could be asked about the other place-holders. But as the specification of the relevant subjects can be included in the specification on the C-conditions, this constitutes no separate problem. With respect to the response, the problem takes a slightly different form; we have already discussed it in the form of Johnston's (1993) worry mentioned in Ch. 4.)

1.1 Strategies for specifying C-conditions

The problem is that specifications of the C-conditions that respect one requirement tend to violate the other. The most common way to specify the C-conditions in the literature on response-dependence is to give long lists of conditions given in (low-level) terms independent of the subject matter under discussion. An example for colours would be this:

- (1) S looks at the object from a distance of between $\frac{1}{2}$ and 10 meters, in lighting conditions similar to those that obtain out of doors on a cloudy Scottish summer day around noon, S pays attention to the object, is competent with the colour concepts, has a visual system that is actually statistically typical for human beings, knows these conditions to be fulfilled etc.²

A first objection to this kind of specification may be that the list consists of descriptions of empirical facts which it takes empirical investigation to know about. This problem is traditionally solved by falling back on a relatively weak notion of a priority, according to which something is a priori if it can be known just in virtue of competence with the concepts. Being competent with the colour concepts requires knowledge of the C-conditions, implicit or explicit, so a competent member of the linguistic community will know these conditions without any further empirical research. Hence they are a priori specifiable.

A more serious problem with the suggested specification is how the 'etc.' is to be filled in. Due to our mastery of colour concepts, we can easily give a rough specification of the C-conditions which, 'suitably belled and whistled', will ensure that judgements and facts about colours co-vary. But there are lots of exceptions that a specification like (1) does not rule out. The question is whether we can 'put the bells and whistles on' such specifications without violating the a priority condition.

In response to this problem, we could replace the 'etc.' with some sort of what-ever-it-takes-to-get-colours-right-clause, in which case the whole exercise will have brought us nowhere; this would make the biconditional a priori, but at the cost of violating the substantiality condition. Alternatively, we could go on adding conditions in order to get a complete list of what is required in order to always get

² This specification is based on Wright (1989), p. 247 (near enough).

correct colour judgements. But this seems equally problematic. First, it is questionable whether the list could ever be completed. For many concepts, it would be practically impossible to circumscribe all error possibilities, and for any candidate specification, there would be no a priori guarantee against further counterexamples. So the list has to be open-ended.³

A second problem is that some of the scenarios to be discounted will probably be rather far-fetched and surprising, and it might take detailed scientific investigation to find out for example what kind of colour illusions arise in which close-to-normal circumstances. If all such scenarios must be taken into account, it is hard to see how the biconditional, including the list of C-conditions, can be a priori true. In short, it seems that the 'long list strategy' will make the equations a posteriori at best, and probably false.⁴

A second strategy for specifying C-conditions would be to give some sort of general, topic-neutral recipe for C-conditions that does not depend on empirical knowledge of colour illusions and their ilk. The problem with such a strategy would be to find an account that gives the C-conditions sufficient concrete content to respect the substantiality condition; in order to do this, the specification must tell us something about what conditions in fact play the role as conditions.

In short, if we choose a substantial, specific formulation of the C-conditions, we get trouble with the a priority condition. If we choose a general, topic-neutral formulation of the C-conditions, it is questionable whether the substantiality condition can be met.

Our dilemma could be rephrased as follows: On the one hand, emphasised by the substantiality condition and the requirement of openendedness, we want the world to have a say in the determination of the C-conditions; it is an empirical question which conditions are favourable for appreciation of e.g. what things are red.

³ Pettit (1999), p. 23-25, states three 'structural' requirements that an account of C-conditions must meet: a priority, non-vacuity (substantiality), and non-closure; in effect, I have restated these requirements, though in a different form. He also states three 'epistemic' requirements about the knowledge that is attributed to members of the linguistic practice (p. 25-27), but we can bracket these for current purposes.

⁴ See Haukioja (forthcoming) for a more elaborate case against this type of specification.

On the other hand, we want the biconditional to be a priori, which means that it should be possible to know just on the basis of knowing how the concepts work that the connection between facts and judgements (/responses /seemings) in C-conditions obtains. It seems that any specification that can give us one of these things fails to provide the other. Is there any way to get both at once?

2. Pettit's functionalist solution

Pettit's 'A Theory of Normal and Ideal Conditions' suggests a way out of this dilemma. Pettit proposes a functionalist theory of C-conditions, where a general, topic-neutral account of C-conditions provides the role specification, and the realisers are conditions like those found on 'long list' specifications like Wright's (1) above. In effect, this suggestion lets the linguistic practice determine the role of the C-conditions, and let the world determine what realisers occupy it.

Pettit gives substance to the functionalist proposal by sketching how this two-tier arrangement comes into being. The story he tells is – unsurprisingly, at this point – the ethocentric story of concept acquisition. Recall that the story is based on the following core elements: First, dispositions to extrapolate from examples in certain ways, based on a salient similarity between instances; secondly, the second order disposition to distrust the deliverances of the first disposition when discrepancies arise between the verdicts of different people or at different times; and third, the disposition to search for factors that might explain such discrepancies – factors which might be responsible for disruption of the normal relationship between facts and responses. (As we have seen, Pettit develops this story with reference to substantial properties in mind, but it works just as well for high-level concepts. A tendency to seek explanations for deviant judgements might be present even if deviant judgements are those that are out of step with a normal regularity pattern, not judgements that are out of step with an underlying substantial property.) Ideally, we want explanations that would make all parties agree on one of the conflicting judgement as the correct one and explain away all the others. One might add that if we can't have that much (e.g. in relativist domains), we aim for as much constancy as we can get.

This story delivers the necessary materials for the functionalist account of C-conditions. The story about the discounting practice puts us in a position to say what the *unfavourable* conditions are; in Pettit's words,

Unfavourable factors will be those factors such that if people were to identify them as perturbances and limitations that undermine detection [of the property in question], then that would maximise expected, long-term convergence among individuals in the use of 'red'. (1999, p. 34.)

With this in place, the C-conditions, or favourable conditions, can be defined as the ones that are not unfavourable – the ones that would be left when unfavourable factors had been discounted. The suggestion could be paraphrased by defining C-conditions as those conditions that would maximise expected, long-term convergence in the use of the concept – or, in the light of Pettit's earlier work, those conditions that would maximise inter- and intrapersonal constancy.

Pettit's story is not one about consensus; it is not that conditions are favourable or unfavourable depending on whether people think of them as such; rather, the way the practice works makes it right to think of some conditions as favourable or unfavourable, independently of what the practitioners think of the matter. Practitioners might well be in error about what unfavourable factors are. Indeed, it might turn out that the whole practice with a concept is based on error; this could happen if there are in fact no conditions that realise the role of C-conditions.

How does this story fare with respect to the challenge of reconciling substantiality and a priority? An initially attractive way to think about the matter is that the role-specification provides the a priority and the realiser level the concrete content necessary for meeting the substantiality requirement. The list of realisers will be a list of concrete, substantial conditions (e.g. being in actually typical daylight conditions, such-and-such a distance from the observer, not in rapid motion relative to the observer, etc.), and this ensures that the substantiality condition is met. (Any attempt to capture it can be left open-ended, as it should be.) This is where the world has its say: empirical features determine which conditions are in fact the ones that play the role of C-conditions as determined by the linguistic practice and the search for convergence in verdicts based on the relevant responses.

(This is not the explanation Pettit offers, however. Instead, he distinguishes two senses in which a statement can be vacuous. There is a weak sense in which every a priori claim is vacuous. And there is the strong sense in which only whatever-it-takes-formulations like those in Blackburn's polar bear example from Ch. 4 are vacuous. Response-dependence biconditionals are not vacuous in the latter sense. On Pettit's view, this is sufficient to meet the substantiality constraint.⁵)

How about the a priority requirement? Pettit's account of this does not focus explicitly on the C-conditions, and does not seem the strongest one that his account would support, so I won't restate it.⁶ A better explanation (in keeping with his earlier work on response-dependence) may be this: For ethocentric concepts with substantial referents, the equations are a priori as a result of the reference-fixing role of responses in C-conditions. In the good cases, the salient similarity response allows us to hook on to a property in the world that normally causes the response in us. Cases in which that causal connection goes astray are discounted in the search for inter- and intrapersonal constancy; they are detected by the irregularities they cause in the normal response-patterns, which surface as inter- and interpersonal discrepancies. In those conditions that would survive the discounting practice, we get things right with regard to the property's extension.

For concepts of response-dependent subject matters, the story would be slightly different: The concept has its extension determined by judgements in best conditions. The conditions that count as best will be those that play a certain role defined by the way the practice with the concepts works. One way to try to capture this role would be in terms of maximal convergence in the use of the concepts. In any case, whatever we say about this role, it is accessible on the basis of considerations about the way the practice with the concepts works. If these considerations are carried out by a person who is competent with the concepts in question, it can be done from the armchair on the basis of her knowledge as a competent concept user, and hence will be a priori.

A functionalist account of C-conditions seems a promising way to reconcile the prima facie conflicting requirements of a priority and substantiality – and maybe the

⁵ Pettit (1999), p. 38-39.

⁶ See Pettit (1999), p. 37-38.

only way. However, as I shall argue below, there are objections, both to the functionalist strategy in general, and to Pettit's particular version of it (or, more precisely, my paraphrase of his views). In the following, I shall explore some of those problems, and some alternative ways of filling out the general strategy.

3. Must both requirements be met by a single specification?

The suggestion outlined above was that a functionalist account of C-conditions allows us to reconcile the a priori and substantiality requirements by letting the role-specification be a priori, while the list of realisers provide the substantiality. This raises the question whether this is sufficient to meet the requirements as originally intended: *Must there be a single specification that meets both constraints?*

If so, advocates of response-dependence must argue either that the substantiality requirement can be met by the role-specification, or that the list of realisers is specifiable a priori. The difficulty of this task was what motivated our problem and the functionalist solution. For the reasons already given, it is hard to see how it could be possible to specify the realiser C-conditions a priori. If Pettit's non-vacuity-interpretation of the substantiality condition is correct, the role-specification could meet this condition (thereby removing part of the motivation for the functionalist account). But if substantiality requires anything more, the prospects would be dim. So if we must require that one specification meets both conditions, the account seems doomed to failure.

This worry is an instance of a more general worry about functionalist theories mentioned in Ch. 3: Do they cheat by pretending to accommodate two 'levels' (and the intuitions that point towards them) at once? Or are they the perfect way to reconcile prima facie conflicting requirements? In the general version, this is a hard question. But in the case of the notion of C-conditions, there might be an answer to hand: we are dealing with a highly theoretical notion, not a familiar phenomenon that we ordinarily take to be one thing (in the same way as we take e.g. colours to be one thing). If there are two (or more) kinds of C-conditions related in a certain way, this need not conflict with any common-sense intuitions. As Pettit has pointed out, the folk need have no notion of C-conditions, as long as they are just discernible in the practice. If two levels of C-conditions are so discernible, this is all well and good.

3.1 How much will be knowable a priori?

The more general questions behind the question considered in the previous section is this: How much can be known a priori about C-conditions? And is this sufficient to satisfy the considerations that motivated the a priority condition?

How much can be known a priori will arguably be different for different versions of response-dependence. For both response-dependence of subject matter and mere ethocentric response-dependence, we should expect a priority for the role-specifications; in order for the theories to make sense, we must have an idea about what sort of conditions the C-conditions are, and this should be discoverable by reflection on the linguistic practice alone.

For the realiser-specifications, it would seem that we have a motivation for a priority some of the way. For concepts of response-dependent subject matters, concept users must presumably have (at least implicit) knowledge of a significant share of the C-conditions, simply because responses in C-conditions determine the extensions of the concepts; if it is totally undecided what cases count, the content of the concept will not be sufficiently determined. This also means that it could not turn out that the C-conditions and the privileged responses were radically different from what we think in the way they could for natural kind concepts; if they were different enough from what we think to substantially change the extension, there would be no justification for saying that we were still talking about the same concept.

In the case of ethocentric concepts with substantial referents, by contrast, there is a substantial property to do the extension-determining. The practice with the concepts must give the concept sufficient content to make it plausible that the term refers to a particular property, which in C-conditions is tracked via the response-patterns. But that is where its job ends. This makes room for the possibility that while our current conception of the C-conditions are sufficient to pick out a referent, we are very far from having the final picture of what conditions are best suited for tracking it, or what detection methods will be the most suitable ones for the property in question. In such a scenario, relatively little of the realiser specification will be accessible a priori. So for ethocentric concepts with substantial referents, there is less of a guarantee of a priori knowledge about the C-conditions than there is for concepts of response-dependent subject matters.

However, even for response-dependent subject matters, the full list of realiser C-conditions will not be specifiable a priori. For example, it could not have been known a priori that rapid motion could generate unusual colour phenomena. (Response-dependence of subject matter is not about consensus only; there is room for underlying properties to do some work (though not as referents), and hence to give rise to surprises like spinning discs.)

Also, there are presumably lower levels of realisers of which nothing can be known a priori. Our examples of realiser C-conditions have been ‘mid-level’ ones: statistically typical daylight conditions, no rapid movement relative to the observer, etc. But these could be translated into claims about photons etc. On this level, nothing would be a priori for either brand of response-dependence.

So the picture of what can be known a priori is a lot more mixed than we might have thought. Viewing the issues through the lens of the functionalist proposal makes this very clear. The next question is whether this is sufficient to satisfy the a priority condition, or at least the considerations motivating the a priority condition.

If we require complete a priori specifiability of the C-conditions, the accounts must fail. But less will do in order to satisfy the considerations that motivated the a priority condition. The a priority condition was motivated exactly by considerations about what should be knowable a priori for response-dependent concepts. Our investigation has shown that when we look at the detail of the C-conditions, we should not expect a priority all the way. This motivates a refinement to the original proposal, not a conclusion that response-dependence accounts must fail because the letter of the original suggestion can’t be met. The demands it is reasonable to make on the basis of the intuitive ideas connected with the various versions of response-dependence are exactly those that would intuitively be fulfilled for concepts that suit those intuitive ideas. So our conclusion should be that it turns out that the a priority condition should not be understood as rigidly as normally assumed.

This conclusion fits well with our conclusion from Ch. 6 that the a priority condition is not the best way to make the distinctions operational. A modified and less clear-cut a priority condition would have been little use for operationality purposes. But if the condition will not serve this purpose anyway, nothing has been

lost. Our conclusion, then, should be that we should require, and expect, a priority to the extent described above, rather than across the board as normally assumed.

A similar discussion could be had about the substantiality requirement, though I'll omit the detail. Some sort of substantiality requirement should certainly be in place – at least Pettit's non-vacuity requirement. But a requirement of full, substantial specifications of roles and realisers may be too strong, as this may be incompatible with the required amount of a priority. A priori but somewhat insubstantial role-specifications and partially a posteriori but substantial realiser-specifications seem a reasonable compromise. If the conditions are given an interpretation that demands more, that interpretation is probably too strong.

4. Constancy for the wrong reasons

Pettit specifies the role of the C-conditions as conditions free of unfavourable factors, i.e. factors the discounting of which would maximise expected, long-term convergence among individuals in the use of the concepts.⁷ I have paraphrased Pettit's specifications as follows: the (role-) C-conditions are those conditions that would maximise inter- and intrapersonal constancy.⁸ In the following, I shall argue that these specifications cannot be right, or at least not complete, and proceed to discuss some alternatives. The problem is that inter- and intrapersonal constancy – and expected long-term convergence – might come about in the wrong ways, i.e. in situations that do not count as favourable conditions. Objections along these lines have been raised independently by Clemens Kappel, Daniel Nolan and Crispin Wright.⁹

⁷ Pettit (1991), p. 34.

⁸ In discussion, he has rejected this paraphrase, so we should be careful with attributing the view to him. On the other hand, it seems a natural paraphrase of his views, and it is not clear whether his account contains the means to avoid the problems to be discussed. I shall phrase the discussion in terms of the constancy formulation, since the objections were made to that formulation, but some of the points carry over to the formulation in terms of 'expected, long-term convergence'.

⁹ Kappel's objection was made at Namicon's workshop on dispositions 08.10.04, Nolan's at the St. Andrews graduate reading party 29.10.04, and Wright's in discussion 17.11.04.

4.1 Deferring to Nolan's sense of humour

Nolan points out that constancy can be *manufactured*. For example, we can obtain perfect inter- and intrapersonal constancy in judgements on humour if we all defer to Nolan on matters of what is funny (provided his sense of humour is stable across time). However, this does not capture the way the concept 'funny' works, so the recipe in terms of maximal inter- and intrapersonal constancy can't be the best way to specify the C-conditions.

A natural rejoinder is that for the purpose of response-dependence views, we are not interested in deferential responses, but in the sort of responses that could ground an ethocentric concept. On Nolan's suggestion, the responses we normally consider relevant to humour will come to play a very limited role. Only Nolan and the few who know they share his sense of humour could rely on them in tracking the relevant (response-dependent) property, while everyone else would need other methods for telling good jokes from bad.

This reply may be problematic. For it might be argued that the concept of Nolan-humour is a perfectly good ethocentric concept; it is based on responses, though only those in a very demanding set of C-conditions that include being Daniel Nolan. Everyone else has to defer to him, but this in itself is no problem, any more than it is a problem that colour blind people have to defer to those with normal colour vision. The best response is probably to concede that 'Nolan-funny' is a response-dependent concept, and argue that it is a different one from the concept 'funny' (for one thing, it has an extra element – being Nolan – in the C-conditions).¹⁰

The same response would apply to the objection that we may get so good at tracking Nolan-humour that we have a 'salient similarity' response of sorts to things that are Nolan-funny. Then an ethocentric concept could develop on the basis of it. But again, it will be a different concept, not the concept 'funny'.¹¹ So Nolan-humour does not pose a serious problem for the view under consideration (nor does it pose a problem to Pettit's original specification).

¹⁰ This point is due to Crispin Wright.

¹¹ Thanks to participants in the Arché relativism seminar for discussion of this point.

4.2 Kappel's taste-good-drug

Kappel's version of the objection is more problematic. Suppose there is a drug that makes people like everything they taste. People under influence of this drug would display perfect inter- and intrapersonal constancy in their judgements concerning matters of taste, and do so on the basis of the same responses as usual. So if the specification of the C-conditions in terms of inter- and intrapersonal constancy was right, the C-conditions for matters of taste should include taking the drug. But taking such a drug would take us far outside the conditions that count as favourable by intuitive lights.

What to do? One option would be to concede that inter- and intrapersonal constancy is insufficient as a role-description of the C-conditions, and try to find supplementary criteria that will rule out the problematic cases. A candidate would be discriminability; in many cases, including colours, it is evident that the conditions that count as favourable are conditions that are favourable for discriminating the finer nuances of the properties in question, besides maximising constancy. Kappel's drug does not maximise discriminability; in fact, it makes everything taste the same. So taking the drug does not qualify as part of the C-conditions.

This won't solve the problem, however. As Wright has pointed out (in discussion), the example might be constructed in such a way that the drug not only maximises constancy, but also enables people to discriminate nuances of taste that are not normally discriminable. So adding a requirement about maximal discriminability to the specification of the role of C-conditions will not solve the problem.

The same problem arises for another response, based on (other) pragmatic considerations. One might argue that responses such as those of taste, and concepts generated on the basis of them, are good for particular purposes, such as navigating in the world and staying alive and well. The drug would make us worse off in such respects. For example, it would make us unable to tell edible from inedible mushrooms of those kinds where poisonousness and bad taste go together. However, this response, too, is vulnerable to the original objection: the drug could be modified to meet these requirements as well.

This sort of amendments to the role-level specification also raises other problems. Adding extra conditions means that the account loses its appealing simplicity, and it raises hard issues about the weighting of the different components. Also, once we have opened the door to such improvements of the role-level specification, it gets hard to close it again. For any choice of role-level specification, it would seem to be an open question whether this specification does in fact comply with the (realiser level) C-conditions at work in the practice.¹²

A better response to Kappel's objection might be to argue that C-conditions are a holistic matter, and that when passing judgement on favourable or unfavourable conditions, we look at our experience across the board, and not only with the properties that the particular concepts refer to. Generally, taking drugs distorts people's conceptions and tends to make them give divergent responses, and so taking drugs that change people's response-patterns from their normal ones cannot count as part of the C-conditions. What we would really like to say about the case is presumably this: Such a drug might be useful, but it guides us to a different property than the one picked out (/constituted) by our normal taste responses. What matters is where the constancy is actually found; not where it would be in far-fetched cases like the drug. We shall return to such ideas below in the discussion of alternative role-descriptions.

However, this response may take us away from the suggestion it was employed to defend, and suggests that C-conditions are a more complicated business than the neat account in terms of inter- and intrapersonal constancy suggests.

4.3 Wright on stable colour illusions

Wright's version of the objection from irrelevant constancy aims to show that cases of constancy outside C-conditions occur naturally, and do not need to be manufactured. He points out that some visual illusions are the same for everyone and at all times, and so provide actual cases of intra- and interpersonal constancy which are classified as falling outside the C-conditions. An example would be the colourful appearance of a spinning black and white disc. This appearance is the same for

¹² I am indebted to Crispin Wright for discussion of this point.

everyone, yet we describe this as a visual illusion rather than a change in colour. Again, the invited conclusion is that a role-specification in terms of constancy (or indeed convergence) will not do.

In response, the holistic line suggested above might be invoked; it is not just that the disc should appear similar to different observers and at different times, but also that the appearance of the disc when spinning should be consistent with its appearance when at rest. What conditions are classified as C-conditions depends on the overall regularity patterns in responses across cases and people, and the colour phenomena generated by spinning the disc is just out of step with the ways colours mostly behave. Hence it is rightly classified as a colour illusion, and as falling outside the C-conditions, and this is done on the basis of considerations about constancy patterns, even if the illusion generated by the spinning disc is the same for everyone. Again, this response may be correct, but also takes us away from the original proposal.

Our discussion suggests that while the functionalist account of C-conditions seems a good and possibly indispensable way to strike the right balance between the a priority and substantiality requirements, the type of role-specification we have considered may be too simple. Unless the holistic story can solve the problems raised without changing the spirit of the proposal too much, maximising inter- and intrapersonal constancy alone will not do. And adding extra clauses like maximising discriminability or being useful for survival is unlikely to help, as such clauses invite counterexamples of the same kind as the original suggestion. The obvious next question is what the role level specification should be if it shouldn't be this.

5. Alternative job descriptions for the C-conditions

How might the role of the C-conditions be described if not in terms of maximising inter- and intrapersonal constancy or 'expected, long-term convergence'? The objections just discussed were fuelled by examples of cases in which the specifications get out of step with the conditions that are accepted as C-conditions in ordinary practice with the concepts. So an obvious place to look for a better role-

specification would be precisely in the conditions accepted as C-conditions in the practice with the concepts.¹³

This suggestion may be filled out in different ways. One version would have it that the C-conditions are those that are *actually, currently recognised as favourable, implicitly or explicitly, in the practice with the concepts*. However, this version faces the objection that the practice with a concept may be on the wrong track, and may treat as conditions as favourable that are in fact a very poor choice. There is such a thing as finding out that what we (all) thought were favourable conditions for a certain kind of judgements are not favourable after all. Relatedly, the practice will often under-determine the C-conditions, leaving open some choices regarding what conditions are favourable. For example, until we discovered the unusual colour phenomena connected with spinning discs, we may not have had a view on whether the conditions favourable for colour judgement excluded rapid motion relative to the observer. Also, there may be cases that are open to decision rather than discovery of which conditions should be taken to reveal the true colours of things. (For example, is the real colours of the television screen those of the pixels or those that appear at normal viewing distance?)

The natural response to these objections is to introduce some sort of idealisation on the conditions actually accepted by participants in the practice – the conditions that would be accepted if we were less ignorant, more rational, etc., or the conditions that will be accepted at the limit of rational enquiry, etc. The problem with this suggestion, of course, is to say exactly what idealisations should be accepted.

The appropriate idealisations may be different for different versions of response-dependence. For response-dependent subject matters, the idealisations imposed on the practice should not take us too far away from the actual practice; since responses in C-conditions are what determines extensions – and arguably content – of these concepts, the motivation for saying that we are still talking about the same concepts will be lost if large variations in the C-conditions are allowed. For

¹³ This approach was suggested by Wright in discussion. Pettit's approach is close to this in spirit; my paraphrases of it may well have taken it further away from the current suggestion than intended by Pettit.

concepts with substantial referents, by contrast, the best conditions for detecting the referent properties may be very different from the ones we currently accept, and so we may be wrong about almost every part of the C-conditions (up to the limit where the concept would no longer pick out the same property as referent). The referents provide a justification for saying that the concept would be the same even if the accepted C-conditions were to change substantially.

A specification of the role of C-conditions as *'the conditions that would be explicitly or implicitly accepted in the practice, suitably idealised'* presumably stands a good chance of avoiding refutation – if for no other reason, then because of its lack of commitment to what a suitable idealisation would be. But this somewhat non-committal formula may be the best we can get if we want something completely sure-fire. No matter how the appropriate idealisations on current practice are specified in more concrete terms, there is presumably a risk of getting out of step with intuitions about what should count as C-conditions, given the way the practice works. Indeed, this is exactly what gets the constancy-based suggestion into trouble.

The suggestion that the C-conditions are the ones that would be accepted explicitly or implicitly in the practice, suitably idealised, is good, indeed obvious. But it would be a shame if this is the final word on the matter. If this is all we can say, the suggestion will have a disagreeable flavour of quietism.

A way around this problem may be to try to say more about what conditions are generally accepted as favourable by the practice and would survive suitable idealisations, and why, without attempting an exhaustive role-specification. The suggestion about inter- and intrapersonal constancy is a good one if understood in this spirit. So are the additions we have considered, such as maximal discriminability and pragmatic value in general.

In particular, the point about constancy seems to capture something crucial to basic concept formation, which even sceptics about response-dependence accept. (In discussions about response-dependence, it often happens that you formulate a response-dependence thesis, and are told that this won't work because the C-conditions can't be specified a priori and substantially. Then you ask, 'But what could basic concept formation be based on if not constancy?', and your interlocutor answers 'of course it must be constancy that grounds basic concept formation'.)

Another suggestion that might serve to elucidate rather than define the (role of the) C-conditions is the suggestion mentioned in Ch. 2: characterising C-conditions as *the conditions in which the relevant features on different levels coincide*. For colours, for example, the suggestion would be that the C-conditions are the conditions in which things that look red also have one of the reflectance spectra and surface structures typical of red things. The idea would be that cases where the normal correlation pattern and causal connection breaks down would be cases in which perturbors are at play, and which should consequently be discounted as unfavourable. (This could be true whether or not the low-level properties qualify as referents of the concepts; the heterogeneity of the physical bases of colours are no obstacle to this characterisation of C-conditions.)

The main problem with this suggestion is how to pick out the ‘relevant features’ in a way that does not presuppose the C-conditions as given. The most straightforward way to identify the ‘relevant features’, in general or on a particular level, is to pick the features that are present in C-conditions. But the combination of these two strategies is blatantly circular. This is why I suggest the account as an elucidation rather than a definition of the role of C-conditions. It may be possible to solve the problem by relying on similarity patterns alone: to pick out the relevant features as those that display certain regularity patterns which are roughly correlated with regularity patterns on neighbouring levels, though they do not coincide in all conditions. But it would take more space than is available here to make the proposal workable, and it is not clear whether it could be sharpened sufficiently to serve to define the role of C-conditions.

To sum up our discussion of alternative role-specifications, we have a general overall recipe that is presumably correct, but not very informative: ‘The C-conditions are the conditions that would be explicitly or implicitly accepted as favourable by the practice given suitable idealisations’. We have considered three ways of filling out this suggestion: The suggestion based on maximising inter- and intrapersonal constancy, the suggestion about the conditions actually accepted by the practice, and the suggestion that the C-conditions are those in which the relevant features on different levels coincide.

What is the relationship between these suggestions? Our discussion has shown that they are indeed different suggestions, and can come apart. We employed some of the differences in arguing against the constancy-based suggestion; the examples of constancy for the wrong reasons are cases in which the constancy-based suggestion comes apart from the conditions actually recognised in the practice. However, there is also reason to expect that the conditions picked out by the different recipes will by and large converge, though I will omit detailed discussion of the connections between the versions.

5.1 Unruly C-conditions

C-conditions are arguably a much more heterogeneous and messy lot than normally assumed in discussions that make use of them. Giving a nice, clean, unified account of them is thus no easy task. In this section, I shall sum up some of the ways in which C-conditions are more unruly than normally assumed.

First, as we shall see in Ch. 8, the C-conditions must presumably be relativised, not only to classes of concepts (e.g. colour concepts), but also to features of the particular situations or cases that the judgements concern. For example, the C-conditions for making colour judgements normally include daylight, but daylight can't be among the C-conditions for judging the colours of e.g. starlight, fireworks and northern lights.

Secondly, C-conditions are probably a more holistic matter than we might have thought. As suggested in the discussion of irrelevant constancy, some factors may be classified as unfavourable for one concept because they cause perturbances relevant to another.

Thirdly, as mentioned above, the C-conditions relevant to a given concept may not be fully determined by the practice with the concepts. The way the practice works will often leave a lot of semantic decisions for later, and will leave room for unexpected discoveries about previously unknown perturbers.

Fourthly (and relatedly), it is probably often left open how the range of normal or favourable cases should be demarcated for a given concept. For example, how wide is the class of competent judges on tastes in wine? We can choose to defer to the few with a very sensitive palate and appropriate training, or to most people, or

some intermediate choice. Likewise with colour vision; we can choose to call many or few types of lighting favourable for colour determination. *Mutatis mutandis* for other parts of the C-conditions, and for the responses themselves in cases where these, too, can be a matter of degree. Often the question about how broad a range of cases count is simply left open, at least for a range of intermediate cases, often because the precise way the limit is drawn has no practical significance (or because there is not enough constancy in responses to draw an exact limit).

C-conditions are thus a far more heterogeneous lot than normally assumed in discussions of response-dependence. If strict a priority and substantiality of a single, simple specification was required for response-dependence accounts to work, then, the accounts would seem doomed to failure for most concepts.

5.2 The consequences for response-dependence accounts

However, less might do. The core idea that Wright wanted to capture in the idea of C-conditions for response-dependent concepts was this:¹⁴ For some concepts, there is no room for the idea of large-scale, inter-subjectively shared error in the cases that we take to be core cases, or the conditions that are most conducive to getting things right with respect to the matter in question. There is simply no higher court of appeal than the judgements made in such conditions – no independently given property that our judgements attempt to track, and which we may get wrong even in the best of conditions. This idea can still make sense even if the class of best judgements turns out to be hard to capture, and less orderly and sharply demarcated than we might have expected.

The main points of Pettit's account of response-dependence and of C-conditions also survive the discovery that C-conditions are hard to capture; it can still be the case that concepts are acquired and evolve on the basis of similarity patterns across cases, observers, and times, even if these similarity patterns are not easily captured in a simple recipe.

The 'under-determination' of the C-conditions fits well with the ethocentric story of concept evolution (and with Wright's (1989) somewhat similar account of

¹⁴ This view was conveyed in discussion.

intentional states). On this story, responses or seemings are trusted until it turns out that something has gone wrong, e.g. because contradictory judgements are encountered across people and times. Only then does the search for explanations in terms of perturbing factors begin. Gradually, knowledge about perturbing and conducive factors can develop (and more semantic decisions be made), and the C-conditions become more and more well-defined. But there is no reason to expect that this process is ever completed for any given concept.

The under-determination of C-conditions also fits well with the idea of level-flexibility – that it can be left open whether a given concept is euthyphronic or has a substantial referent. The C-conditions may be expected to develop in different ways in the two cases, and if they had to be given a final form before the concept could function, this would presumably entail that the location question would have to be settled in advance. But the fact that the C-conditions are not completely worked out from the beginning might explain how the practice with both kinds of concepts might work in roughly the same way before the location question is settled, even though the final story about the determination and detail of the C-conditions will be very different for concepts with low-level referents and concepts of response-dependent subject matters.

Finally, the ‘under-determination’ in the C-conditions might explain how it can be that the different suggestions about role-specifications we have discussed all sound sensible, even if they can come apart in some cases; these cases, or most of them, will presumably be cases that we haven’t yet had reason to think about.

6. Conclusion

In response to the problem of reconciling a priority and substantiality in the specification of the C-conditions, the best approach seems to be this: First, we should adopt a functionalist account that combines a role-specification which can be given a priori – near enough, and with provision made for strange- and fool’s cases in the case of Pettit-style response-dependence – with a realiser-specification that provides the substantiality, but will not be knowable a priori (for ethocentric concepts with response-independent referents), or only parts of which can be known a priori (for concepts of response-dependent subject matters). The best candidate for a role-level

specification which is available a priori is the very unspecific ‘the conditions that would be implicitly or explicitly accepted by the practice given suitable idealisations’. Filling this suggestion out in more concrete terms is risky business, as we saw with the Pettit-based suggestion in terms of constancy; for any such specification, counterexamples might be forthcoming. However, suggestions like this have an important role to play as elucidations of general features of the C-conditions, even if they won’t work as definitions. The suggestions discussed are interesting even if viewed as part of such a more modest project. In particular, Pettit’s point about constancy seems to capture a feature that is crucial to basic concept formation.

A second lesson of our discussion is that we should recognise the limitations of the requirements traditionally associated with response-dependence proposals. We should not expect a priority across the board in the specification of the C-conditions, and the fact that it’s not available is no objection to response-dependence accounts. Nor should we expect C-conditions to be too well behaved. They may be under-determined by the practice, relative to particular cases, etc. Thus, a simple, general account of them may not be forthcoming. In this sense, the accounts may be less substantial than we might have expected. Again, this is no serious obstacle to response-dependence theses. It raises problems only on a too narrow conception of what is needed to get such theses to work.

If we were relying on a priori biconditionals fulfilling certain conditions in order to make the distinctions operational, it would be a problem if the C-conditions could not be specified in a more precise way than this. It would be hard to distinguish response-dependent from response-independent concepts because the C-conditions would leave a lot of scope for correction and decision and, presumably, for response-independent concepts to pose as response-dependent. But given that we don’t, the under-determination, relativity, holism, and elusiveness of C-conditions does not have to threaten the proposals.

The conditions Wright employed to capture the distinction were designed to ensure response-dependence. But in many cases (particularly the independence condition), the conditions were ways of ‘playing it safe’, and setting aside some hard questions for later consideration; the conditions were designed to ensure response-dependence, but not to do so in the least demanding way. Thus, a discovery that

some of the conditions are too strong in the way they have traditionally been understood is no disaster for the idea of response-dependence. When the justification of the demands is thought through, it may turn out that less will do.

There is lots of further work to do on all these issues. But I think there is reason to expect that the results of such work will vindicate the idea(s) of response-dependence, rather than proving it flawed.

Ch. 8: Conditional fallacy problems

Response-dependence theses seem vulnerable to conditional fallacy problems like those that affect the simple conditional analysis of dispositions. This chapter is an attempt to clear response-dependence these of those charges. I discuss what the counterexamples show, and how they might be resisted. I consider four attempted solutions to the problems: Johnston's appeal to dispositional formulations, Wright's provisional equations, Blackburn's 'elasticity' approach, and a strategy based on relativisation of the C-conditions. I conclude that some of these suggestions offer the resources to solve the problem.

1. Conditional fallacy problems

Conditional fallacy problems are best known for the counterexamples they pose to the conditional analysis of dispositions. I shall use the standard examples from this context to present the problems.¹

To say that something has a disposition is to say that it can be expected to react in a certain way to certain stimuli or situations. If something is fragile, it will normally break when struck; if something is poisonous, it will cause harm if ingested; if someone is brave, she will act appropriately in dangerous situations. Knowledge of such patterns of stimuli and responses is useful for predicting events and choosing actions; this is why disposition ascriptions are practically important. Accordingly, a conditional analysis of dispositions has seemed a natural choice:

- (1) A system x has the disposition d at time $t \leftrightarrow$ if x were to be subjected to stimulus s at t , it would give response r ²

¹ See Shope (1978) for a survey of parallel problems in many other areas. Shope coined the commonly used but misleading label 'conditional fallacy problems' (misleading because no *fallacy* need be involved; but since this label is standard vocabulary, I shall use it anyway).

² This version of the conditional analysis does not work for all dispositions. For example, it doesn't apply to dispositions to elicit certain responses in other beings and things, or dispositions where no particular stimulus is specifiable. But we can ignore these complications for current purposes.

To take a specific example, we can focus on this (simplified) analysis of fragility:

(2) A vase is fragile at $t \leftrightarrow$ if the vase were struck at t , it would break

The conditional analysis of dispositions seems vulnerable to four types of counterexamples: those based on finkish dispositions, finkish lacks of dispositions, masked dispositions, and mimicked dispositions. A finkish disposition is a disposition that is lost when the characteristic stimulus occurs so the response is never manifested. A frequently used example is a sorcerer who has decided to protect his favourite fragile vase by making it solid if it was struck. As long as the vase is not struck, it remains fragile; the analysandum of (2) is true. But were it to be struck, it would become solid, and so would not break; hence the analysans of (2) is false.³

A finkish lack of a disposition is the complementary case: x lacks the disposition, but would gain it if s occurred, and would do so in time to respond with r . This yields cases of false analysandum and true analysans. An example would be a stone vase which a bad-tempered sorcerer has decided to make fragile if it is struck. While unstruck, the vase is solid, so the analysandum of (2) is false. But thanks to the sorcerer, it would break if struck; the analysans is true.

In masking and mimicking cases, the disposition itself does not change as it does in finkish cases, but the response is prevented or provoked by external factors. In masking cases, the object has the disposition, and the stimulus occurs, but external factors prevent the response. An everyday example would be a vase packed in polystyrene to protect it from breaking if accidentally struck during transport. As with finkish dispositions, such cases yield true analysandum, but false analysans. Mimicking is the complementary phenomenon: the object lacks a given disposition, but if the stimulus characteristic of that disposition were to occur, the characteristic response would be brought about by external factors. An example would be a solid stone vase filled with nitro-glycerine that would explode and break the vase if it were struck. In such cases (as with finkish lacks), the analysandum is false, and the analysans true.⁴

³ A non-supernatural example – that of a circuit breaker – is found in Martin (1994).

⁴ Mark Johnston was the first to discuss masking and mimicking cases in print (Johnston (1993), appendix 2, p. 119-121, and (1992), p. 231-34 (an extension of the passage in Johnston 1993).

These problems for the conditional analysis of dispositions do not have a generally recognised solution. David Lewis (1997a) has suggested an amended conditional analysis, supplemented with a reference to the causal bases of the dispositions. His analysis is widely believed to solve the problems with finkish dispositions and finkish lacks. But it is as vulnerable to masking counterexamples as the simple conditional analysis (as argued by Bird (1998) and Choi (2003)). There are other interesting suggestions on the table, but none are generally accepted.⁵

In response to the problems with the conditional analysis, some argue that dispositions should be identified with their bases, and that the project of giving an analysis should be given up (e.g. Bird 1998). A related strategy would be eliminativism, the view that disposition talk should be viewed as a confused way to talk about bases, and should be given up. Others (including Martin 1994) conclude that dispositions are primitive *sui generis* properties, and that the conditional analysis fails not because of the conditional element, but because it is an analysis. Yet others (Gundersen 2002) argue that the counterexamples are merely apparent, and so do not threaten the conditional analysis of dispositions after all.

1.1 Conditional fallacy problems for response-dependence theses

Response-dependence theses formulated in terms of subjunctive conditionals give rise to problems parallel to those that afflict the conditional analysis of dispositions. Assume for the sake of the argument a simple response-dependence account of colours:

(3) x has colour $F \leftrightarrow x$ would look F to standard observers in standard conditions

Suppose x is a piece of photo-sensitive paper, waiting to be used for making photographs in a darkened lab. We would normally say that the paper is white;

Martin (1994; publication delayed for more than a decade according to Lewis (1997a)) is often credited with the introduction of finkish cases, though awareness of conditional fallacy problems goes further back; see e.g. Shope (1978). Johnston and Wright prefer the term ‘altering’ to ‘finkish dispositions’ and ‘finkish lacks’, but the idea is the same. I shall stick to the latter terminology which is more widely used.

⁵ See e.g. Gundersen 2002 and Mumford (1996).

presumably it has a surface structure which would usually look white in C-conditions, and in the soft red lighting of the lab, it looks the same colour as the lab assistant's white T-shirt and coffee mug. Indeed, if it was any other colour, it would presumably be useless for making photographs. However, if the paper was taken outside and placed in normal daylight, it would look black, as the light would change its surface structure. So the right-hand-side of the biconditional is false. Yet intuitively the paper is white as long as it remains securely in the lab, so the left-hand-side is true. This looks like a counterexample to the biconditional.

Another standard example is Johnston's shy but intuitive chameleon.⁶ This creature is uniformly green as it sits on a leaf in the dark (or, in less question-begging terms, it has a surface structure normally found in objects that look green). It is also shy and very perceptive; if a being with a suitable visual system were to spot it, it would immediately notice and blush bright red. We would like to say that the chameleon is green (and disposed to turn red). But the right hand side of (3) is false for $F=\text{green}$ and true for $F=\text{red}$, even as the chameleon slumbers peacefully on its leaf. So we get counterexamples to (3); for red, the left-hand-side is false and the right-hand-side true, and conversely for green.

These counterexamples are of the 'finkish' type where x 's coming into C-conditions alters x in such a way that it would no longer elicit the response. What about masking and mimicking counterexamples, in which x itself does not change, but where the response is prevented or provoked by external factors? Such cases do not seem to pose a specific problem for response-dependence theses in the way finkish cases do. They can be avoided if the C-conditions are specified in such a way that no masker or mimicker would be present. This might be hard to accomplish, but it shouldn't be harder than ruling other perturbing factors that might make facts about F and judgements about F come apart. Finks, by contrast, do their work by changing the object when it *enters* C-conditions, and so cannot be dealt with by specifying the C-conditions in a way that rules them out. Finkish counterexamples exploit a contrast between how things are outside C-conditions and how they are in C-conditions, and the possibility of such a contrast can't be undermined by working on the C-

⁶ Johnston (1992), p. 231 and (1993), p. 119.

conditions alone. But with masking and mimicking, the situation is different: if they can be ruled out in the specifications of the C-conditions, the problems will be solved. Hence we shall not discuss masking and mimicking further in this chapter.

The response-dependence thesis in (3) is formulated in terms of subjunctive conditionals. What happens if, instead, the response-dependence thesis is formulated in dispositional terms, e.g.

(4) x has colour $F \leftrightarrow x$ is *disposed to* look F to standard observers in standard conditions?

Do such formulations give rise to conditional fallacy problems as well? The answer to this question depends on how we think of dispositions. If dispositions are to be given a simple conditional analysis, then (4) will be equivalent to (3); to say that x is *disposed to* look F to standard observers in standard conditions is just to say that it *would* look F to standard observers in standard conditions. In that case, of course, the two formulations will give rise to exactly the same problems. On this view, response-dependent concepts (/properties) will be a subclass of dispositional concepts (/properties), and conditional fallacy problems for response-dependence theses will be instances of the general problem for dispositions.

If we give a different account of dispositions, the situation is more complicated. If an alternative account of dispositions can be found that avoids conditional fallacy problems, this account may perhaps be carried over to response-dependence theses and provide a solution to the conditional fallacy problems apparently afflicting the original proposals. Then that account, whatever it is, might form the basis of fink-free response-dependence claims. This is where Johnston places his hopes. On the other hand, the account of dispositions may be such that it won't help us to capture an interesting response-dependence distinction. If the account involves a low level location of dispositions, then it might change the spirit of dispositional response-dependence theses beyond recognition, and so may not provide a way out. Likewise, an eliminativist account of dispositions won't be any help in saving response-dependence theses from conditional fallacy problems.

1.2 What is under pressure from the objections?

If the conditional fallacy problems can't be overcome, how much damage will be done to the response-dependence proposals? In the case of the conditional analysis of dispositions, many have taken the problems to be knock-down objections. Some might want to draw the same conclusion in the case of response-dependence accounts. But it doesn't have to be this bad. First, it might be possible to show that something is wrong with the counterexamples, and so that the equations are left undamaged. Secondly, even if there is a problem, it may be a problem with the equations rather than with the intuitive ideas the equations are employed to capture. The intuitive ideas will be undamaged if the equations can be altered slightly to avoid the problems while still doing their work. Thirdly, the ideas may even survive the discovery that no simple formula captures them completely, provided that the conditional fallacy problems can't be rephrased in a way that strikes directly at the intuitive content.

2. Attempted solutions

I shall argue that the literature on the topic offers ways to solve the problems. I shall consider four different strategies. Two are versions of the 'counterexamples are only apparent'-strategy: Blackburn's (/Gundersen's) strategy of arguing that the counterexamples are results of inconsistent applications of intuitions on the two sides of the biconditional, and a strategy based on relativising the C-conditions. If these two strategies are combined, they yield a promising suggestion. The two others are instances of the strategy of adjusting the equations to avoid counterexamples. One is Johnston's strategy of formulating response-dependence in overtly dispositional terms, and appeal to a better understanding of dispositions for a solution to the problems. The other is Wright's strategy of replacing simple biconditionals with 'provisional equations'. I also briefly mention, but do not develop or discuss, a strategy based on Lewis's suggestion regarding dispositions: that an appeal to bases should be built into the biconditionals.

2.1 Johnston on response-dependence and conditional fallacy problems

One of the very visible differences between Wright's and Johnston's accounts of response-dependence is that Johnston's is phrased in terms of dispositions, while Wright uses subjunctive conditionals. Ironically, both motivate their choice by saying that it, as opposed to the other, can deal with conditional fallacy problems.

Johnston takes conditional fallacy problems to refute both the simple conditional analysis of dispositions and response-dependence accounts based on subjunctive conditionals. He thinks the problems should be solved by finding a better account of dispositions that avoids such counterexamples. Once such an account is found, it can be employed in formulating response-dependence theses that are not vulnerable to conditional fallacy problems. Until the right account is found, the dispositional idiom can function as a place-holder for the correct account of dispositions that we don't yet know. This, Johnston thinks, allows us to bracket the problems until we have an account of dispositions that avoids them.

Should we agree in his optimism? I think not. First, the approach seems unsatisfactory because Johnston straightforwardly helps himself to an account of dispositions we don't yet have. Unless it is supported by reasons to believe that an account of dispositions is forthcoming that would solve the problems, the account contains an element of wishful thinking. It would seem strange to claim this as an advantage over views that actually try to address the problem.

A more serious problem is that the correct theory of dispositions, once discovered, may not fulfil Johnston's expectations. Not all possible accounts of dispositions would suit Johnston's purposes. If the correct account of dispositions turns out to be one that identifies them with their bases, it is hard to see how dispositions could form a basis for response-dependence accounts for the reasons stated earlier. A high level location of dispositions would be a better basis for a theory of response-dependence. But if a high-level location is chosen, it may be that nothing better than the conditional analysis is forthcoming. If so, an account formulated in terms of dispositions would have no advantage over views formulated in terms of subjunctive conditionals. An eliminativist theory of dispositions would, of course, be an even less welcome conclusion for Johnston's purposes. A primitivist account would seem to be the option best suited for the task (and fits well with many

of Johnston's claims about dispositions). But even such a view may not provide the perfect basis for response-dependence theories. On a primitivist view, there may not be much to say about dispositions, and consequently an account of response-dependence formulated in dispositional terms might be correspondingly uninformative.

In short, there is no guarantee that a complete understanding of the nature of dispositions, if we were to gain one, would be one that would fulfil Johnston's expectations and provide a basis for response-dependence theses immune to conditional fallacy problems. So Johnston's strategy is unsatisfactory.

2.2 Wright on response-dependence and conditional fallacy problems

Wright's approach to the problem is very different from Johnston's. Given the apparent failure of the conditional analysis of dispositions and the lack of a better theory of dispositions, he thinks it inappropriate to formulate response-dependence equations in dispositional terms. His suggested solution to conditional fallacy problems is to replace 'basic equations' of the form

$$(5) \quad F(x) \leftrightarrow (C \rightarrow S \text{ would judge that } F(x))$$

with 'provisional equations' where the C-conditions are placed in a proviso:

$$(6) \quad C \rightarrow (F(x) \leftrightarrow S \text{ would judge that } F(x))$$

Basic equations are vulnerable to conditional fallacy problems, since the C-conditions' coming to obtain could alter the truth value of Fx . But provisional equations say only what would happen in cases where the C-conditions are already met, so any changes in x induced by their coming to obtain will already have taken place in the relevant cases, and so will not give rise to counterexamples. This move makes the equations immune to 'finks' on the C-conditions.

Finks on the response?

This move seems to solve the problem with finks triggered by the object's coming into C-conditions. But it also raises problems. One possible problem is that the equations may be vulnerable to conditional fallacy problems that arise elsewhere. If the change in x 's colour is triggered by the *response* rather than by the C-conditions

coming to obtain, counterexamples may still arise. An example might be a very shy and very perceptive person who will notice any glance in her direction and instantly blush.

Wright's rejoinder to this objection (given in discussion) is, first, that it takes more than a shy teenager to make a counterexample. It takes a telepath, since she must be able to perceive the exact moment at which the other person makes a *judgement* regarding her complexion, and not only whether she is being watched. But he adds that this response is somewhat unsatisfactory, since we would want the account to work even for far-fetched or supernatural cases. A better response is that in order to work, the argument requires that judgements in C-conditions involve a two-stage process: that first, the C-conditions are met, and then a judgement is made. This presupposition can be challenged. If a case can be made that both must come about at once, the problem will be solved. One way to do this would be to build into the C-conditions a requirement that a judgement is made. Whether such a requirement is legitimate depends on how the C-conditions work; if not, a clause to the effect that a judgement is made may simply be added in the proviso in conjunction with the C-conditions. This move might seem ad hoc, but it would presumably solve the problem.

The univocity objection

A second and more serious challenge to the provisional equation strategy is *the univocity objection*. Unlike basic equations, provisional equations do not say what is true of e.g. colours in general, but only what is true of colours *provided that the C-conditions are fulfilled*. In principle, this opens the possibility that extensions are determined in two (or more) very different ways: in C-conditions, they are determined by best opinions, but outside C-conditions they could be determined by something quite different, e.g. underlying physical properties. This possibility is an unwelcome consequence. It suggests an ambiguity in the concepts which intuitively isn't there.⁷

Wright offers two responses to this challenge. The first is that the objection underestimates what it takes to generate an ambiguity. A genuine ambiguity requires

⁷ Wright raises and discusses the objection in (1992), p. 125-127.

the concept to have two or more different meanings, for which different explanations are required. But response-dependence equations are not in the business of offering such explanations at all. Indeed, the ‘echo’ in the equations – the fact that the terms under discussion appear on both sides of the biconditional – makes them useless as explanations. Their job is merely to explicate an a priori feature of the concept.

Wright’s second point against the challenge concerns colour concepts in particular. It is that colours supervene on physical characteristics (surface structures, reflectance spectra, or whatever), and that once responses in C-conditions have paired the colour concepts with a certain group of physical characteristics, the latter can be employed to ‘extend’ the concept to cases that fall outside C-conditions. (This does not require a simple physical common factor, but is consistent with heterogeneity on the underlying levels.)

Wright’s answers make the objection seem less than fatal to the account. But it may still be objected that the shift to provisional equations cause a loss of generality of the resulting response-dependence accounts, and that an account that could do without the restriction imposed would be more appealing and easier to work with.

It might be argued, however, that this objection – and the univocity objection itself – hinges on a conception of the equations as doing more than response-dependence theorists intend them to do. The critic seems to want an all-round recipe of how to apply the concept, and objects that if the equations concern only applications in C-conditions, it is an entirely open question how the concept should be used in other cases. But this would only be relevant if the job of the equations was to prescribe the application conditions of the concept. And this is not how they are intended.

The claim of response-dependence proposals (of subject matter) is that the concepts have their extensions determined by best opinions. As mentioned, the core thought behind Wright’s version of response-dependence was that there are concepts for which there is no higher court of appeal than judgements in the most favourable conditions. Outside these conditions, we try to find the best approximation to applications in C-conditions. But what that approximation is may be left partly open by the proposals, and may be determined in different ways for different concepts or even cases of the same concepts. It may be that no general recipe is forthcoming.

Provisional equations seem well equipped to capture this situation, neutral as they are on what happens outside C-conditions. ‘

The objection gets one thing right: there is a good question about how the extension of a response-dependent concept is determined outside C-conditions. What makes this a good question is that applications outside C-conditions may be determined in various ways even for the same concept; by physical similarities, or by appearance alone, or by appearance corrected for known perturbing factors. But this need not damage the proposal; in fact, it might strengthen it. For response-dependent concepts, the important thing is that F’s are the things that seem F in C-conditions, and that is exactly what is captured in provisional equations. Provisional equations are therefore a good way of stating response-dependence theses, and also provide a promising solution to the problem with finkish cases.

3. Blackburn’s ‘elasticity’ approach

Blackburn (1993) suggests another interesting approach to the problems. The suggestion is that the alleged counterexamples hinge on elasticities in the way we think about cases like Johnston’s chameleon. If looked at in one way – holding fixed the surface structure and imagine looking – the chameleon has one colour: green. If looked at in another way – holding fixed its entire nature, including its tendency to blush, and imagine looking – it has another: red. Both are legitimate ways to imagine what the chameleon would look like – or, in Blackburn’s preferred version of the biconditional, what colour experience it ‘is naturally such as to elicit’.⁸ Blackburn does not make explicit how it helps to distinguish these two ways of looking at the cases. But the thought may be that both options, if applied consistently across the biconditionals, give the right results: If you think of the chameleon in the first way, it would look green (or, in Blackburn’s terminology, is naturally such as to look green), and indeed intuitively it is green. If you think of the chameleon in the second way, it would (/is naturally such as to) look red, and sure enough, intuitions would have it that it is red. Trouble arises only if the two ways of thinking about the case are

⁸ Blackburn 1993, p. 265.

mixed, so one perspective is employed on one side of the biconditional and another on the other.⁹

A natural way of filling out Blackburn's suggestion would be to argue that the alleged counterexamples hinge on level equivocations. (This is probably extension rather than exegesis; some passages of Blackburn's suggest that this might be his view, but most suggest that it isn't.) The thought would be that if colours are located on the level of physical properties, we should think of the chameleon as green, whereas if they are given a high-level location and thought of as essentially connected to the response-pattern – if performance in C-conditions is all that counts – then we should think of the chameleon as red. If we hesitate to say that the chameleon is red even before the light goes on, this could be ascribed to the physicalist intuitions that are arguably part of colour concepts. Hesitation to admit that the chameleon is green though it would look red in C-conditions, on the other hand, suggests that we think of colours as high-level, response-dependent properties. The argument might continue as follows (from here definitely not blameable on Blackburn): If either line of thought is consistently applied to the cases under consideration, the problems disappear. If colours are located with surface structures, then the chameleon in the dark is green, and it doesn't matter that it may appear differently in C-conditions. If intuitions tell us otherwise, they should be corrected. Conversely, if we think of colours as high-level properties, we should admit that the creature is red even in the dark, as this is what it would look like in C-conditions. The intuitive pull towards saying that it is green stems from physicalist intuitions which should be given up if we opt for a high level location. Either way, purifying the concepts of the equivocation would solve the problem. Or so the thought goes.

But unfortunately, purifying the concept of physicalist intuitions doesn't solve the problem for response-dependence theses. For the claim that the photo paper has a different colour in the lab than in daylight can be justified by appeal to appearances alone, and without drawing on physicalist intuitions. The main motivation for calling

⁹ This way of stating the diagnosis, especially the last sentence, is not based on evidence in Blackburn's text, but is inspired by Gundersen (2000, 2002).

the paper white is that in the soft, red lab light, it looks the same pink colour as things (coffee mugs etc.) that we know to look white in daylight. (We might describe the case as a case that is outside C-conditions, but where we know how to correct for the perturbing influence of the conditions – i.e. to think of pink-looking things as white). Also, if the paper was subjected to daylight and then taken back into the lab, it would be useless for making photographs, and would now appear more similar in colour to the coffee than to the mug. This suggests that daylight changes the paper, even if we don't think about surface structures or similar. So even if you think that colours as essentially appearance-related, you might justifiably think that the paper is white in the lab. The temptation to think so does not disappear if we purify the concepts of physicalist intuitions. So blaming the problem on a level equivocation is no solution.

3.1 Beyond the level contrast

The contrast Blackburn is after can also be described in a way that does depend on a level-equivocation (and that was probably Blackburn's intention). The two ways of viewing the disputed cases may be thought of as two ways of viewing the physical situation, or two ways of viewing the relevant appearances, rather than a physicalist and a response-dependence-based way of viewing the situation. If you think about the surface configuration of the chameleon as it is and imagine seeing it, the result will be that the chameleon is green (since that's the way it would look, one might add; this story does not prejudge the location issue). But if you hold fixed the physical nature of the object as a whole (i.e. a broader system that includes the finkish mechanism), we get the opposite result: the chameleon is such as to look (and would look, and is disposed to look) red.

(This story contains the seeds of the level contrast explored above; if you want a physical location for the colours, it would be reasonable to look for properties that are frequently present in white things – the surface structures – rather than with other properties of the objects, such as those responsible for the finkish mechanisms. Conversely, if the second way of thinking sounds convincing to you and you're willing to think of the paper as black just because that's what it would look like in

daylight, then that suggests a high level location. But even if we bracket the location questions, the contrast is present.)

Blackburn concludes that finkish problems can be avoided by awareness of the elasticities in concepts that allow the contrasts to arise. Viewed in one way, chameleons and their ilk have one colour, and would make true the corresponding right-hand-side, and viewed in another way, they have another, and would make true the right-hand-side for that colour.¹⁰

However, there is a problem with this attractive line of thought. The problem is that intuitions about the colour cases seem unevenly distributed across the biconditionals. If the 'either way' strategy was to work, intuitions should have it that to the extent that the photo paper is white, it would look so in C-conditions, and to the extent that it is not, it should not. But this is not the result we get. There is a lot of intuitive pull towards thinking of the photo paper as white: its appearance in the lab and its uselessness for photo production once it has been subjected to daylight. Furthermore, there is a strong intuition that the paper actually changes on coming into C-conditions. Such changes are not just stipulation, but a phenomenon from everyday life: you can watch the photo paper in various conditions, and when the light becomes too strong, you watch it change. An acceptable response-dependence thesis of colours should respect this intuition and have the paper come out white before the change, and black afterwards.

But intuitions about the right hand side of the biconditionals favour the conclusion that the paper is black: colours are normally ascribed to objects, and the object as a whole contains a finkish mechanism, and hence it is such as to look black in daylight conditions. (If you think that only the surface structure needs to be taken

¹⁰ Blackburn 1993, p. 267.

A more elaborate proposal of similar spirit is found in Gundersen (2000) and (2002). Gundersen argues that the alleged counterexamples all fail because they are impaled on one or other horn of a dilemma. On one horn, they work by illicitly strengthening the antecedent of the subjunctive conditionals. On the other, the object, or system, is chosen differently on the two sides of the biconditional. Either way, the 'counterexamples' fail. Unfortunately, presenting and discussing Gundersen's suggestion in the depth it deserves would take too much space to be feasible in this context.

into consideration, consider a case where the fink is part of the surface structure.) So the problem is not straightforwardly solved by saying that for each of two different ways of looking at the matter, the biconditional will come out true, and that problems arise only by illicit combinations of left- and right-hand-sides.

You might think that the adjustment of intuitions needed to solve this problem is so small that it should be permissible, if the reward is a solution to the problem with ‘finkish’ counterexamples. However, this may not sound convincing to all, especially not to people who are sceptical about response-dependence. So as stated, the suggestion is not an attractive solution to the problems.

4. Relativising C-conditions?

The photo paper is intuitively white in the lab, but it would blacken if taken into daylight. This is a counterexample to the response-dependence biconditionals if, but only if, the C-conditions for judging the colour of the photo paper include daylight. Could this be the assumption we should challenge in order to solve the problem? If the C-conditions for determining the colour of the paper do not involve daylight, but only lab light, we avoid the conclusion that the paper in the lab would look black in C-conditions. In discussions of response-dependence, it is often assumed without argument that the C-conditions are relative to classes of disputed concepts, but uniform for all concepts and all cases within such a class. But this seems too crude. The optimal conditions for judging the colours of e.g. fireworks, starlight, and northern lights clearly isn’t daylight, even though daylight forms part of the C-conditions for most colour determination tasks. So the assumption of uniform C-conditions for all colour determination should be given up in any case.

For the photo paper, too, there is a case for saying that the optimal conditions for determination of its colour is lab light, not daylight. The case is admittedly less clear than the other examples. The special C-conditions would apply only for some colours the paper might have, not others, and would no longer apply when the photo-making process is completed, or when the paper has been ruined by daylight. But still it seems clear enough for the suggestion to be taken seriously.

4.1 Some unsuccessful objections

A 'relativist' view of C-conditions as the one proposed might raise a range of worries. Let's start by heading off some of them. First, if it is so obvious that C-conditions should be relativised to cases in the way proposed, how come this has been missed? So far, everybody has treated C-conditions as uniform across entire domains. The answer is probably that the notion of C-conditions is relatively new and hasn't been explored much yet, and also that the response-dependence accounts that make use of it raise many other issues worth working on. So it is no wonder if some of the complexities have gone unnoticed or been put aside for later consideration. Also, we should note that the consensus about domain-wide C-conditions is not based on arguments; it's just the way things have been done this far. If we have good reasons for dissatisfaction with it, this consensus is not enough of a reason to resist further relativisation.

Next, is the invocation of relativised C-conditions as a solution to the finkish problems an ad hoc move? The answer is probably no. For if we look at the candidates we have for general accounts of the C-conditions (role level descriptions), they all seem to agree on relativisation to cases. If we were after fully specified lists of C-conditions, of course, relativisation to (smaller classes of) concepts and cases would seem hopeless, as it would make the already daunting task much larger. But the fully specified list approach is hopeless anyway, and so its hopelessness for this purpose should not count against relativisation of C-conditions. Indeed, the idea of relativised C-conditions suggests an explanation of why the requirement of a fully specified list is unjustified as well as unnecessary: the practitioners can be relied upon to contextualise aptly, though they may not be able to articulate the rationales in each case.¹¹

We have discussed three better candidates for general recipes of C-conditions. One is the Pettit-style, constancy-based account (maybe with additions on discriminability etc.). Another is the 'overlap of relevant factors on different levels' account, and the third is the 'C-conditions are those accepted in the practice' approach. Each of these not only seem compatible with relativisation of C-

¹¹ Thanks to Sarah Broadie for suggesting this.

conditions; they seem to actually recommend it. Maximal constancy and discriminability in judgements on e.g. northern lights would be reachable in the dark, not in daylight. The overlap between visual impressions and the complex physical bases of the phenomenon would also be found in such conditions. Finally, people generally accept that the night is the time to look for northern lights; no-one waits for the daylight in order to see them better. Thus, all the most likely candidates for a general recipe of C-conditions would entail that the C-conditions for appreciation of e.g. fireworks and ordinary surface colours are different – in other words, they would dictate relativised C-conditions.

Thirdly, is the relativity to cases compatible with the substantiality requirement common to most response-dependence accounts? The answer is yes. We can't say whatever we like about what the relativised C-conditions are, any more than we can for other realiser C-conditions. Relativisation raises no new problems in this respect, and does not generate triviality.

So, should we conclude that the apparent conditional fallacy problems are caused by too simplistic thinking about the C-conditions? As usual, things are not quite that simple.

4.2 A serious objection

A seemingly serious problem is that relativising the C-conditions does not rule out all finkish cases. The problem is the same as the one that came up regarding the level equivocation approach discussed earlier in the chapter: appearances outside C-conditions count, too. In the photo paper example, the lighting conditions in the lab may have a fair claim to be the C-conditions for the case concerned. But in other cases, the situation is different. If we take a case that is firmly outside C-conditions, but where the object appears F, and where this is not undermined by the perturbing factors at play, but where an unusual change in appearance would follow if the object was put into C-conditions, we get the same problem as before. The chameleon example, if altered slightly, is a case in point. If we suppose that it is possible to get a glimpse of the creature even in the dark on the rare occasions when it is asleep and doesn't notice the observer, we get a problem case: the chameleon would look greenish in the dark – or in any case indistinguishable from the leaf, even though the

lack of light would not allow us to see its colour clearly. But if we turn on the light, it would start looking red. And the relativisation of C-conditions move will not help; it *is* easier to judge the colour of the creature once the light is on.

Could the problem be solved by further relativisation? The suggestion would be that *whenever there is a colour change involved, the C-conditions can't be assumed to be the same*. Thus, if the chameleon is such as to change its colour on coming into daylight, there may be – and is – a different set of C-conditions for judging its colour outside daylight conditions. When the colour changes, the C-conditions change with it. This squares well with intuitions: if you want to know the colour of the photo paper in the lab or the chameleon in the dark, you'd better find a way of investigating that doesn't cause colour changes. So you look for the best conditions for determining the colour compatible with not changing it.

4.3 An independence worry

This suggestion invites an objection based on considerations similar to those that motivate Wright's independence condition: if the C-conditions are defined relative to colour changes, how can responses in C-conditions be what determine the extensions of colour concepts? The appeal to colour changes in picking out the C-conditions looks like a serious case of presupposing extensions of the concepts as already given in the specification of the C-conditions.

There are two ways to respond to this worry, both controversial. One suggestion is that perhaps a more holistic and less exact story could be told. The claim would be that C-conditions work in a more complex way than assumed in early work on response-dependence, and that C-conditions and extensions are determined *together*, rather than one being in place before the other.

The changes in question might be described in different ways. If we describe them as changes in distribution of colour properties, it sounds very much like begging questions or presupposing independently given extensions. But if we describe them as sudden changes in appearance in a certain familiar respect as conditions are varied, and where this change is unexpected in the light of the normal patterns of changes as a function of changing conditions, the story starts sounding less problematic. Something like this may be sufficiently different from explicit

presuppositions about extensions to work. If the idea is that C-conditions settle where we get the right amount of constancy (and/ or co-instantiation with lower level common factors), then either such changes in appearances are permissible determinants of the C-conditions, or nothing is.

The complex way C-conditions seem to work in ordinary linguistic practice gives reason for optimism about the relativisation strategy, as does the vagueness with which they seem to be defined. As suggested in Ch. 7, C-conditions are a confusing and under-defined lot. For example, the ‘wideness’ of the class of cases to be taken as C-conditions is often left open: should we defer on matters of taste to those selected few with a very sensitive palate? Or should we include the judgements of most of us, only discounting those with a very weird taste or who have just brushed their teeth? In real life, we clearly do both. Recognising the complexities involved may provide the means to defend a solution to conditional fallacy problems based on relativised C-conditions.

(However, even if this strategy is correct, it may be badly placed to convince those who are negatively inclined towards response-dependence. It will presumably take a charitable approach to see the mileage of a proposal of this kind.)

The other response to the independence worry advertised takes a completely different approach: To justify the claim of change of C-conditions by appealing to changes in the bases of the dispositions – or, better, *intuitions about changes of bases*. The idea would be that as long as the colours are different from the bases, the bases would give us a way of describing the changes that justify claims about changes of C-conditions without describing them as changes in colour.

This suggestion may sound heretical in the context of defending response-dependence theses – theses that locate the reference of the concepts on a high level rather than on the level of bases. However, appeals to bases are presumably compatible with response-dependence of subject matter as long as no *identification* with bases is made. No such identification is in question here; indeed, it would undermine the purpose of appealing to bases in the first place.

Note that if the strategy of relativising C-conditions works, it would also offer a way of defending Blackburn’s suggestion about the objection about the asymmetry of intuitions about the two sides of the biconditional. If the C-conditions we consider

are lab light for the white photo-paper and daylight for black photo-paper, we get the right intuitive results in both cases. This combination might seem to be unnecessary if the relativisation strategy can do the work alone. However, Blackburn's strategy also offers important insights, which might make the combination preferable over the relativisation strategy alone.¹²

The strategy based on relativisation of C-conditions calls for elaboration, and for further work on defending it against the objections discussed. But it seems a promising approach, alone or in combination with the Blackburn/ Gundersen strategy.

4.4 A base-based solution to conditional fallacy problems?

Appeals to low-level correlates or bases of response-dependent properties have come up several times in our discussion: in Wright's response to the univocity objection, in the level-equivocation reading of Blackburn's strategy, and in the defence of the relativisation strategy against the independence worry. The question naturally arises whether conditional fallacy problems for response-dependence theses could be addressed in a more direct way by appeal to changes of bases. A suggestion along these lines would be to adopt Lewis's strategy from 'Finkish Dispositions', where he defends a (revised) conditional analysis of dispositions. The move that solves the problems with finkish dispositions and lacks is the introduction of base-properties into the equation. Lewis's analysis is this:

- (7) Something x is disposed at time t to give response r to stimulus s iff, for some intrinsic property B that x has at t , for some time t' after t , if x were to undergo

¹² This also applies to Gundersen's more elaborate proposal. When presented with the relativisation strategy and the challenge to his and Blackburn's views based on intuitions being unevenly distributed across the biconditionals mentioned above, he responded by suggesting that the problems should be solved by choosing the cases differently: the two cases to be contrasted are simply photo paper in lab light and photo paper in daylight. This combination seems very attractive, though I don't have the space to show its potential.

stimulus *s* at time *t* and retain property *B* until *t'*, *s* and *x*'s having of *B* would jointly be an *x*-complete cause of *x*'s giving response *r*,¹³

where *B* is the causal base of the disposition, and an *x*-complete cause is a complete cause as far as *x*'s intrinsic nature is concerned. The right hand side of this analysis is true even for finkish cases, because even if in those cases *x* loses the disposition and its base and so does not respond with *r*, it is still true that if it had retained *B* long enough, it would have responded in the appropriate way. The problem with finkish lacks is solved by requiring that *B* is a property that *x* already has, and not one it gains as a result of the stimulus. If a similar move was incorporated in response-dependence biconditionals, this would presumably solve the problems with finkish cases.

Again, such appeals to bases might sound heretical in the context of response-dependence. But the suggestion would be, not to identify the response-dependent properties with their bases, but only to connect them with claims that *there is a base property* that is doing a certain job. The account would be neutral on what sort of properties do the work, whether they are substantial properties or combinations of factors, and whether the base is the same from case to case. The appeal to bases would entail that the relevant response-dependent properties supervene on physical properties (probably in a fairly strong sense), but not that they are identical with them.

Such a proposal would raise several hard problems, and would take a lot of discussion to be made good. Unfortunately, that discussion would take us across the word limit, so it will have to be omitted. But it is a strategy that deserves investigation, and such investigation would be a natural part of an investigation into the relation between response-dependent properties and their low-level correlates – an investigation which, again, we cannot take on here, but which would be a natural next step in the investigation begun by this project.¹⁴

¹³ Lewis (1997a), p. 157.

¹⁴ Even if the strategy turns out not to work for response-dependence of subject matter, it might work against similar problems for Pettit-style response-dependence where substantial referents are in the picture. For ethocentric concepts with low-level referents, it would be natural to seek a solution by appeal to low-level properties.

5. Conclusion

'Finkish' cases like photo-sensitive paper and Johnston's chameleon pose a prima facie problem for response-dependence theses. But the response strategies we have considered contain the resources for solving it. The best response may be to follow Wright and formulate response-dependence claims in terms of provisional equations. These seem very well suited to capture the core idea behind (at least Wright's version of) response-dependence proposals: that there is no higher court of appeal than responses in best conditions, and that these determine the extension of the concept. Applications in less favourable circumstances are presumably made by best approximations to what happens in C-conditions; these approximations can be made in various ways, but this is no objection to the proposals.

An alternative suggestion is provided by Blackburn's elasticity approach, though as it stands, this strategy seems insufficient to solve the problem because of the objection from asymmetric distribution of intuitions across the biconditionals. A strategy based on relativising C-conditions seems more promising, though this, too, faces serious objections, and the suggested answers to these are not beyond questioning. But if the suggestion can be made to work, it offers another solution to the problem, and a more nuanced understanding of C-conditions into the bargain. It might be combined with Blackburn's or Gundersen's strategy to yield an even more nuanced proposal. Finally, we have touched on the possibility of solving the problems by appealing to bases, applying a move from Lewis's revised conditional analysis of dispositions. This strategy would be worth exploring, especially in the context of investigating the relation between response-dependent properties and their low-level correlates.

In all, we can conclude that conditional fallacy problems do not pose a serious threat to response-dependence theses.

Conclusion

In this section, I sum up the results of the investigation and point to some of the issues left for further investigation. An important if obvious result is that response-dependence comes in very different versions. The main divide is between theories like Wright's and Johnston's that locate the disputed phenomenon on a high level, and thus make the very subject matter under discussion response-dependent, and theories like Pettit's that concern the concepts only and are neutral on the location issue. Within these general paradigms, further distinctions can be made. We have distinguished two routes to response-dependence theories of subject matter: 1) an a priori route from intuitions of e.g. transparency to the conclusion that the concept is criterially governed with responses in C-conditions as main criterion, and 2) an a posteriori route relevant for level-flexible concepts that turn out to lack appropriate low-level referents.

Furthermore, the relevant high-level locations can be described in lots of different ways. Wright describes them in terms of an Euthyphronic order of determination, where best opinions determine the extensions of the concepts. Johnston describes them in terms of response-dispositions. And there are more ways to do it, e.g. Wedgwood's description in terms of response-dependent essences. I have argued that some of the differences between the various ways of describing response-dependence theses of subject matter are not as significant as they might have seemed. In particular, dispositional and subjunctive formulations will be equivalent if dispositions are thought of as high-level properties (as seems appropriate for the purposes of response-dependence theses). Also, while the relevant high-level locations might be described in terms of reference to slightly substantial dispositional properties or in terms of criterially governed concept correlated with merely deflationary 'referents', this contrast may be viewed as an artefact of two ways of looking at the same matter (i.e. a third-person and a first-person approach to dispositions).

Thus far, our results are largely compatible with the various accounts of response-dependence of subject matter. But when it comes to the best way of making the distinctions operational, our conclusions have been somewhat revolutionary. I

have argued that the traditional apparatus of biconditionals or provisoed biconditionals are not the best way to make the distinctions operational, but that the focus should be on ‘test cases’ like the possibility of strange- and fool’s cases, the possibility of reference failure, and the hints about location provided by causal intuitions, unity intuitions and transparency intuitions. Furthermore, I have argued that the possibility of strange- and fool’s cases mean that the traditional biconditionals will not be a priori for Pettit-style response-dependence.

For Pettit-style response-dependence, the results have been a similar blend of positive and negative lessons. Pettit’s ethocentric story has very interesting potential, including a suggested solution to the Wittgensteinian rule-following problem, and the answers to our ‘hard questions’ from Ch. 5. However, the most interesting version of the story is not Pettit’s version about the concept acquisition of individuals, but a story about the concept acquisition of entire communities. This version – the ethocentric story of concept evolution – can explain how concepts can exist and be in good standing before the nature of their referents, or even the levels of the referents, is known. The story explains how concepts come into being on the basis of response-dispositions in subjects that make certain cases seem saliently similar, and a discounting practice aimed at inter- and intrapersonal constancy in verdicts based on the ‘salient similarity’ responses, which delivers the C-conditions and a seems/is distinction. This story is neutral on the location issue, and allows it to remain open on what level the referents of the concepts should best be located. Yet it also allows low-level properties to have their say in determining the extensions of the concepts, and thus explains how concepts made of ‘for us’ material could come to refer to response-independent properties. Furthermore, the story suggests a unified account of how concepts as different as natural kind concepts, concepts of response-dependent subject matters, and concepts of relativist subject matters, might come into being; all these types of concepts could evolve in the way described by the ethocentric story.

The various versions of response-dependence we have distinguished are not competitors for a single job, but supplement each other well (in particular, the ethocentric story provides a plausible account of how concepts of response-

dependent subject matters might come into being). So a project of finding THE correct version of response-dependence in general would be misconceived. The situation is rather the opposite: operating with system of different distinctions enables us to respect the differences between the subject matters to which we apply it, and to capture nuances that would be lost if only one distinction was in play. The combination of Pettit-response-dependence and response-dependence of subject matter allows us to distinguish three overall ways response-dependent concepts might function: they might be high-level-rigid, or level-flexible with high-level referents, or level-flexible with low-level referents.

For a particular domain, finding the appropriate version(s) of response-dependence would be a worthwhile project, of course. Unfortunately, we have not been able to go into detail about any particular location disputes, though a little has been said about colours, intentional states, meaning, and dispositions. But it is presumably in such discussions that the response-dependence distinctions, and in particular the levels framework, will really show their worth (or lack of same).

None of the objections to response-dependence that we have considered seem insurmountable, though many deserve more attention than it has been possible to give them here. In particular, this is true of Johnston's missing explanation argument, which I had hoped to discuss in more detail.

The *prima facie* conflict discussed in Ch. 7 between the a priority and substantiality conditions can be resolved by adopting a functionalist account of C-conditions. For this suggestion to solve the problem, the conditions must be understood in a less rigid way than normally assumed, but this is no problem because the considerations that motivate them can be satisfied by less demanding interpretations. Nor is it a serious problem that C-conditions turn out to be a much more heterogeneous and unruly lot than ordinarily assumed; response-dependence theses can still work, even if they can't be captured in as simple accounts as we might have thought.

The conditional fallacy problems discussed in Ch. 8 can also be addressed. The most straightforward solution is to shift the focus to provisoed biconditionals; other possibilities are relativisation of the C-conditions (possibly in combination with the

Blackburn/ Gundersen strategy), or perhaps shifting the focus to Lewis-style biconditionals that mention the bases of the relevant response-dispositions.

Our discussion has raised a lot of issues that call for further investigation. Further work would be useful regarding the exact way to formulate, and to distinguish between, the various versions of response-dependence. Also, there is work to do regarding the notion of C-conditions, the work it does for response-dependence theses, and whether a better characterisation of C-conditions can be found than the ones we have considered.

One of the big questions left open by our investigation concerns the relationship between response-dependent properties and their lower-level correlates. I suspect that an investigation into this matter would provide lots of interesting results. I also suspect that the notion of supervenience would be helpful in such an investigation, since it provides scope for variation (regarding the ‘broadness’ of supervenience bases, the relevant worlds, etc.), and hence might respect, and help us get a grip on, the differences between different domains. An investigation into these matters would be a natural next step in the investigation begun in this work. Another topic that deserves further work is the levels framework. I have developed it enough in Ch. 2 and 3 to serve our purposes, but there is lots of further work to do regarding all the questions discussed in Ch. 3 (level-individuation, the idea of a basic level, etc.).

The levels framework has proved very helpful in the discussion of response-dependence. First, it has helped us distinguish the various versions and gain a clear perspective on their similarities and differences. Secondly, it has helped to bring out the potential of the accounts. Most notably, the levels framework plays a crucial role in the considerations about how an ethocentric story of concept evolution answers the ‘hard questions’ discussed in Ch. 5. It also paves the way for arguments for response-dependence theses, e.g. by exposing prejudices like the idea of a basic level, and also by suggesting a way to think about properties which allows for substantiality in degrees, rather than viewing substantiality as a matter of all or nothing. Perhaps the most important advantage of the levels framework in the context of response-dependence is its usefulness for setting up the location disputes

in which response-dependence theses are candidate options, though our discussion has been too abstract and general to demonstrate this advantage clearly.

Successful applications like this is probably the best justification we could give for endorsing the framework. The brief discussion of material constitution in the appendix to Ch. 3 suggested another area in which we might expect the idea of levels to do useful work.

The discussion has suggested that response-dependence is good for a lot of things. However, in a certain respect, the results in the light of the levels framework might be disappointing. We might have hoped that response-dependence theses could offer the perfect compromise between intuitions about realism or objectivity, on the one hand, and intuitions about subject-relatedness and epistemic access on the other. We might have hoped that response-dependence theses would allow their subject matter to be subject-related (and hence accessible to minds like ours), yet completely real and objective. The result has been a part-way vindication: response-dependence, even response-dependence of subject matter, does not entail anti-realism in the senses of non-factualism, unreality, or epistemic inaccessibility; only the question about subject-relatedness has to be answered in a non-realist way. But response-dependence of subject matter entails that the phenomenon is not a low-level one, with all that follows. This will unavoidably contradict some realist intuitions, even if the levels framework makes it possible to view high level properties as real.

In this work, I have tried to take the idea – or ideas – of response-dependence as seriously as possible. It has probably been evident that I am positively inclined towards the basic ideas, and expect quite a lot from response-dependence, especially the ethocentric story, and the system of distinctions that has emerged. I do not claim that the proposals are unproblematic, nor that response-dependence theses can be defended for every domain and at all costs. There are still lots of difficult problems to handle, and a lot of work to do in precisifying and applying the different distinctions in the area. But I think there is reason for optimism. It seems fair to conclude that the idea of response-dependence, in its different guises, makes sense.

Bibliography

- Ayer, Alfred J. (1936): *Language, Truth and Logic*, Gollancz, London
- Bird, Alexander (1998): 'Dispositional Antidotes', *Philosophical Quarterly* 48, p. 227-34
- Blackburn, Simon (1984): *Spreading the word*, Clarendon Press (OUP), Oxford
- Blackburn, Simon (1993): 'Circles, Finks, Smells and Biconditionals', *Philosophical Perspectives* 7, Language and Logic, p. 259-280
- Block, Ned (1995): 'Can the Mind Change The World?', in Cynthia Macdonald (ed.): *Philosophy of Psychology: Debates on Psychological Explanation*, Blackwell, Cambridge
- Block, Ned (2003): 'Do Causal Powers Drain Away?', *Philosophy and Phenomenological Research* 67, p. 133-50
- Boghossian, Paul (1989): 'The Rule-Following Considerations', *Mind* 98, p. 507-49
- Boghossian, P. & D. Velleman (1989): 'Colour as a Secondary Quality', *Mind* 98, p. 81-104
- Byrne, A. and D. R. Hilbert (eds.) (1997a): *Readings on Color, Vol. 1: The Philosophy of Color*, MIT Press, Cambridge, Massachusetts
- Byrne, A. and D. R. Hilbert (eds.) (1997b): *Readings on Color, Vol. 2: The Science of Color*, MIT Press, Cambridge, Massachusetts
- Campbell, J. (1993): 'A Simple View of Colour', in Haldane and Wright (ed.): *Reality, Representation and Projection*, Oxford University Press, Oxford
- Casati, R & C. Tappolet (eds.) (1998): *Response-Dependence, European Review of Philosophy* 3
- Chalmers, David (1996): *The Conscious Mind*, Oxford University Press, Oxford
- Choi, Sungho (2003): 'Improving Bird's Antidotes', *Australasian Journal of Philosophy* 81 (4), p. 573-80
- Crane, Tim (2001): *Elements of Mind*, Oxford University Press, Oxford
- DeClercq, Rafael (2002): 'Two Conceptions of Response-Dependence', *Philosophical Studies* 107, p. 159-77
- Dennett, Daniel (1988): 'Quining Qualia', in A. Marcel & E. Bisiach (eds.): *Consciousness in Contemporary Science*, Oxford University Press, Oxford

- 1988; Reprinted in D. Chalmers (ed.): *Philosophy of Mind: Classical and contemporary readings*, Oxford University Press, Oxford 2002
- Devitt, Michael (2006): 'Worldmaking Made Hard: Rejecting Global Response Dependency', *Croatian Journal of Philosophy* 6 (16), p. 3-26
- Divers, John & Alexander Miller (1999): 'Arithmetical Platonism: Reliability and Judgement-Dependence', *Philosophical Studies* 95, p. 277-310
- Edwards, Jim (1992): 'Secondary Qualities and the A Priori', *Mind* 101, p. 263-72
- Field, Hartry (1989): *Realism, Mathematics and Modality*, Blackwell, Oxford
- Galilei, Galileo (1957): 'The Assayer' (1623), in *Discoveries and Opinions of Galileo*, Doubleday Anchor, New York
- Geach, Peter (1980): *Reference and Generality*, Cornell University Press, Ithaca
- Gundersen, Lars Bo (2000): 'Bird on Dispositional Antidotes', *Philosophical Quarterly* 50 (199), p. 227-29
- Gundersen, Lars Bo (2002): 'In Defence of the Conditional Account of Dispositions', *Synthese* 130 (3), p. 389-411
- Gundersen, Lars Bo (2004): 'Outline of a new semantics for counterfactuals', *Pacific Philosophical Quarterly* 85 (1), p. 1-20
- Hardin, C. L. (1988): *Color for Philosophers: Unweaving the Rainbow*, Hackett, Indianapolis
- Hare, R. M. (1952): *The Language of Morals*, Oxford University Press, Oxford
- Hare, R. M. (1991): 'Universal Prescriptivism', in P. Singer (ed.): *A companion to Ethics*, Blackwell, Oxford
- Haukioja, Jussi (2000): *Rule-Following, Response-Dependence and Realism*, Reports from the Department of Philosophy, University of Turku (Ph.D. thesis), Turku
- Haukioja, Jussi (2001): 'The Modal Status of Basic Equations', *Philosophical Studies* 104, p. 115-22
- Haukioja, Jussi (2005): 'Hindriks on Rule-Following', *Philosophical Studies* 126, p. 219-39
- Haukioja, Jussi (2006): 'Why the New Missing Explanation Argument Fails, Too', *Erkenntnis* 64 (2), p. 169-75

- Haukioja, Jussi (forthcoming): 'How (not) to specify normal conditions for response-dependent concepts', forthcoming in *Australasian Journal of Philosophy* 2007
- Hindriks, Frank (2005): *Rules and Institutions – Essays on Meaning, Speech Acts and Social Ontology*, Erasmus University Rotterdam (Ph.D. thesis), Rotterdam
- Hobbes, Thomas (1985): *Leviathan*, Penguin Books, London (first published 1651)
- Holton, Richard (1991): 'Response-Dependence and Infallibility', *Analysis* 51, p. 180-84
- Holton, Richard (1993): 'Intention Detecting', *Philosophical Quarterly* 43 (172), p. 298-318
- Hume, David (1978): *A Treatise of Human Nature*, (eds. L. A. Selby-Bigge & P. H. Nidditch), Clarendon Press, Oxford (first published 1739-40)
- Hume, David (1975): 'An Enquiry Concerning Human Understanding', in *Hume's Enquiries*, (ed. L. A. Selby-Bigge & P. H. Nidditch), Clarendon Press, Oxford (first published 1748)
- Hütteman, Andreas & David Papineau (2005): 'Physicalism Decomposed', in *Analysis* 65 (1), p 33-39
- Jackson, Frank (1986): 'What Mary didn't know', *Journal of Philosophy* 83, p. 291-95
- Jackson, Frank (1994): 'Finding the Mind in the Natural World', In R. Casati, B. Smith & G. White (eds.): *Philosophy and the Cognitive Sciences*, Holder-Pichler-Tempsky, Vienna. Reprinted in D. Chalmers (ed.): *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, Oxford, 2002
- Jackson, Frank (1998): *From Metaphysics to Ethics*, Oxford University Press, Oxford
- Jackson, Frank & Philip Pettit (2002): 'Response-dependence Without Tears', *Philosophical Issues* 12, *Realism and Relativism*, p. 97-117
- Johnston, Mark (1989): 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society* Supplementary vol. 63, p. 139-74

- Johnston, Mark (1991): 'Explanation, Response-dependence and Judgement-dependence', in Menzies (ed.) *Response-dependent Concepts*, ANU Working Papers in Philosophy 1, p. 122-183
- Johnston, Mark (1992): 'How to Speak of the Colours', *Philosophical Studies* 68, p. 221-63
- Johnston, Mark (1993): 'Objectivity Refigured: Pragmatism Without Verificationism', in Haldane and Wright (eds.): *Reality, Representation and Projection*, Oxford University Press, Oxford, p. 85-133
- Johnston, Mark (1998): 'Are Manifest Qualities Response-Dependent?', *The Monist* 81 (1), p. 3-44
- Kim, Jaegwon (1993): 'The non-reductivist's troubles with mental causation', in *Supervenience and Mind*, Cambridge University Press, Cambridge 1993, p. 336-58 (pagination in quotes are from this edition); reprinted from John Heil & Alfred Mele (eds.): *Mental Causation*, Oxford University Press, Oxford, 1993, 189-210
- Kim, Jaegwon (1998): *Mind in a Physical World*, MIT Press, Cambridge, Massachusetts
- Kim, Jaegwon (2003): 'Blocking Causal Drainage and Other Maintenance Chores with Mental Causation', *Philosophy and Phenomenological Research* 67 (1), p. 151-76
- Kripke, Saul (1972): *Naming and Necessity*, Harvard University Press, Cambridge, Massachusetts
- Kripke, Saul (1982): *Wittgenstein on Rules and Private Language*, Basil Blackwell, Oxford
- Kvanvig, Jonathan (1999): 'Lewis on Finkish Dispositions', *Philosophy and Phenomenological Research* 49 (3), p. 703-10
- Lebar, Mark (2005): 'Three Dogmas of Response-dependence', *Philosophical Studies* 123, p.175-211
- Lewis, David (1973): 'Causation', in Lewis: *Philosophical Papers II*, Oxford University Press, Oxford 1986
- Lewis, David (1989): 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society* Suppl. vol. 63, p. 113-138

- Lewis, David (1997a): 'Finkish Dispositions', *Philosophical Quarterly* 47, p.143-58
- Lewis, David (1997b): 'Naming the Colours', *Australasian Journal of Philosophy* 75 (3), p. 325-43
- Locke, John (1951): *An Essay concerning Human Understanding*, Everyman, London (first published 1690)
- Lopez de Sa, Dan (2003): *Response-Dependencies: Colors and Values*, University of Barcelona (Ph.D.-thesis; accessible on the author's webpage)
- Lopez De Sa, Dan (2006): 'Rigid vs. Flexible Response-Dependent Properties' (unpublished manuscript)
- Mackie, J. L. (1977): *Ethics: Inventing Right and Wrong*, Harmondsworth, Penguin, London
- Malzkorn, Wolfgang (2000): 'Realism, Functionalism and the Conditional Analysis of Dispositions', *Philosophical Quarterly* 50 (201), p. 452-69
- Martin, Charles B. (1994): 'Dispositions and Conditionals', *Philosophical Quarterly* 44 (174), p. 1-18
- McDowell, John (1981): 'Non-Cognitivism and Rule-Following', in Holtzman and Leich (eds.): *Wittgenstein: To Follow a Rule*, Routledge, London, p. 141-62
- McDowell, John (1985): 'Values and Secondary Qualities', in Honderich (ed.): *Morality and Objectivity: A Tribute to John Mackie*, Routledge, London, p. 110-29
- McFarland, Duncan (1999): 'Mark Johnston's Substitution Principle: A New Counterexample?', *Philosophy and Phenomenological Research* 59, p. 683-89
- Menzies, Peter (ed.) (1991): *Response-dependent Concepts*, ANU Working Papers in Philosophy 1
- Menzies, Peter (1998): 'Possibility and Conceivability: A Response-Dependent Account of Their Connections', in *European Review of Philosophy* 3, p. 255-77
- Miller, Alexander (1989): 'An Objection to Wright's Treatment of Intention', *Analysis* 49 (4), p. 169-73

- Miller, Alexander (1995): 'Objectivity Disfigured: Mark Johnston's Missing-Explanation Argument', *Philosophy and Phenomenological Research* 55, p. 857-68
- Miller, Alexander (1997): 'More Responses to the Missing-Explanation Argument', *Philosophia* 25, p. 331-349
- Miller, Alexander (1998): 'Rule-Following, Response-Dependence, and McDowell's Debate with Anti-Realism', *European Review of Philosophy* 3, p. 175-97
- Miller, Alexander (2001): 'The Missing-Explanation Argument Revisited', *Analysis* 61, p. 76-86
- Miscevic, Nenad (1997): 'Secondary and Tertiary Qualities: Semantics and Response-dependence', *Pacific Philosophical Quarterly* 78, p. 363-79
- Miscevic, Nenad (1998): 'The Aposteriority of Response-Dependence', *The Monist* 81 (1), p. 69-84
- Mumford, Stephen (1996): 'Conditionals, Functional Essences and Martin on Dispositions', *Philosophical Quarterly* 46, p. 86-92
- Mumford, Stephen (1998): *Dispositions*, Oxford University Press, Oxford
- Nolan, Daniel (2005): *David Lewis*, McGill-Queen's University Press, Montreal & Kingston; Ithaca
- Pettit, Philip (1990): 'The Reality of Rule-Following', *Mind* 99, p. 1-21
- Pettit, Philip (1991): 'Realism and Response Dependence', *Mind* 100, p. 587-626
- Pettit, Philip (1998a): 'Noumenalism and Response-Dependence', *The Monist* 81 (1), p. 112-33
- Pettit, Philip (1998b): 'Terms, Things and Response-Dependence', *European Review of Philosophy* 3, p. 61-73
- Pettit, Philip (1999): 'A Theory of Normal and Ideal Conditions', *Philosophical Studies* 96, p. 21-44
- Pettit, Philip (2005): *Rules, Reasons, and Norms*, Oxford University Press, Oxford
- Pettit, Philip (2006): 'Response-dependent Theories', *Encyclopaedia of Philosophy* (2nd edition), Borchert Macmillan Reference, New York
- Pettit, Philip and Peter Menzies (1993): 'Found: the Missing Explanation', *Analysis* 53, p. 100-109
- Plato (1992(!)): 'Euthyfron', in *Platons Skrifter* 1, Hans Reitzels Forlag, Copenhagen

- Powell, Mark (1998): 'Realism or Response-Dependence?', *European Review of Philosophy* 3, p. 1-14
- Price, Huw (1998): 'Two Paths to Pragmatism II', *European Review of Philosophy* 3, p. 109-48
- Prosser, Simon (2006): 'Temporal Metaphysics in Z-land', *Synthese* 149, p. 77-96
- Putnam, Hilary & Paul Oppenheim (1958): 'Unity of Science as a Working Hypothesis', in H. Feigl, M. Scriven and G. Maxwell (eds.): *Concepts, Theories, and the Mind-Body Problem*, Minnesota Studies in the Philosophy of Science 2, p. 3-36
- Putnam, Hilary (1967): 'The Nature of Mental States', in Rosenthal (ed.): *The Nature of Mind*, p. 197-204, Oxford University Press, Oxford 1991
- Putnam, Hilary (1992): 'Why Functionalism Didn't Work', in H. Putnam: *Words and Life*, Harvard University Press, Cambridge, Massachusetts, 1994
- Rea, Michael (1997): 'Introduction', in M. Rea (ed.): *Material Constitution*, Rowman & Littlefield, Lanham, p. xv-lvii
- Schaffer, Jonathan (2003): 'Is there a Fundamental Level?', *Nous* 37 (7), p. 498-517
- Shope, Robert K. (1978): 'The Conditional Fallacy in Contemporary Philosophy', *Journal of Philosophy* 75 (8), p. 397-414
- Smith, Barry (ed.) (1998): *Secondary Qualities Generalized*, *The Monist* 81
- Smith, Michael (1989): 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society* Suppl. vol. 63, p. 89-112
- Smith, Michael (1994): *The Moral Problem*, Blackwell, Oxford
- Smith, Michael (1998): 'Response-Dependence Without Reduction', in *European Review of Philosophy* 3, p. 85-108
- Stalnaker, R. (2004): 'Conceptual Truth and Metaphysical Necessity', *Philosophical Studies* 18
- Tye, Michael (2002): 'Visual Qualia and Visual Content Revisited', in D. Chalmers (ed.): *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press, Oxford
- Vallentyne, Peter (1996): 'Response-dependence, Rigidification and Objectivity', *Erkenntnis* 44 (1), p. 101-112

- Wedgwood, Ralph (1998): 'The Essence of Response-Dependence', in *European Review of Philosophy* 3, p. 31-54
- Wiggins, David (1987): 'A Sensible Subjectivism', in his *Needs, Values, Truth*, Blackwell, Oxford
- Wittgenstein, Ludvig (1953): *Philosophical Investigations*, Blackwell, Oxford
- Wittgenstein, Ludvig (1956): *Remarks on the Foundation of Mathematics*, Blackwell, Oxford
- Wright, Crispin (1986): 'Rule-following, Meaning and Constructivism', in C. Travis (ed.): *Meaning and Interpretation*, Blackwell, Oxford, reprinted in C. Wright: *Rails to Infinity*, Harvard University Press, Cambridge, Massachusetts, 2001
- Wright, Crispin (1988): 'Moral Values, Projection and Secondary Qualities', *Proceedings of the Aristotelian Society* Supplementary volume 62, p. 1-26
- Wright, Crispin (1989): 'Wittgenstein's Rule Following Considerations and the Central Project of Theoretical Linguistics', in A. George (ed.): *Reflections on Chomsky*, Oxford University Press, Oxford, p. 233-265
- Wright, Crispin (1992): *Truth and Objectivity*, Harvard University Press, Cambridge, Massachusetts
- Wright, Crispin (1998): 'Euthyphronism and the Physicality of Colour: A Comment on Mark Powell's 'Realism or Response-dependence?''', *European Review of Philosophy* 3, p. 15-30
- Wright, Crispin (2001): *Rails to Infinity*, Harvard University Press, Cambridge, Massachusetts
- Wright, Crispin (2002): 'The Conceivability of Naturalism', in Gendler and Hawthorne (eds.): *Conceivability and Possibility*, Oxford University Press, Oxford, p. 401-41
- Wright, Crispin (2006): 'Intuitionism, Realism, Relativism and Rhubarb', in P. Greenough & M. Lynch (eds.): *Truth and Realism*, Oxford University Press, Oxford, p. 38-60
- Yablo, Stephen (1987): 'Identity, Essence, and Indiscernibility', *Journal of Philosophy* 84, p. 293-314; reprinted in J. Kim & E. Sosa (eds.): *Metaphysics, An Anthology*, Blackwell, Oxford 1999

- Zangwill, Nick (1994): 'Moral Mind-Independence', *Australasian Journal of Philosophy* 72 (2), p. 205-18
- Zangwill, Nick (2001): 'Skin Deep or In the Eye of the Beholder?': The Metaphysics of Aesthetic and Sensory Properties, *Philosophy and Phenomenological Research*, 61 (3), p. 595-618
- Zangwill, Nick (2003): 'Against Moral Response-dependence', *Erkenntnis* 59, p. 285-90