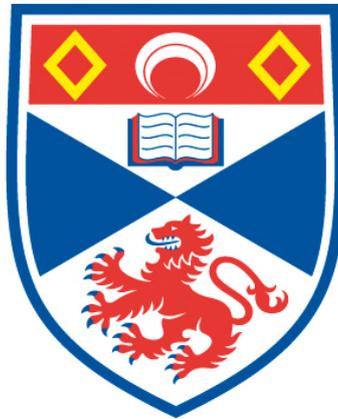


# DISAGREEMENT AND PHILOSOPHICAL METHOD

James Cook

A Thesis Submitted for the Degree of MPhil  
at the  
University of St Andrews



2015

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

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

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/9458>

This item is protected by original copyright

# Disagreement and Philosophical Method

James Cook



University of  
St Andrews

This thesis is submitted in partial fulfilment for the degree of

MPhil at the

University of St Andrews

Date of Submission

26/03/2015

**1. Candidate's declarations:**

I, James Cook, hereby certify that this thesis, which is approximately 35710 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2013 and as a candidate for the degree of MPhil in September, 2012; the higher study for which this is a record was carried out in the University of St Andrews between 2012 and 2015.

*(If you received assistance in writing from anyone other than your supervisor/s):*

I, ....., received assistance in the writing of this thesis in respect of [language, grammar, spelling or syntax], which was provided by .....

Date ..... signature of candidate .....

**2. Supervisor's declaration:**

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

Date ..... signature of supervisor .....

**3. Permission for publication:** *(to be signed by both candidate and supervisor)*

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

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

PRINTED COPY

- a) No embargo on print copy
- b) ~~Embargo on all or part of print copy for a period of ... years (maximum five) on the following ground(s):~~
  - ~~Publication would be commercially damaging to the researcher, or to the supervisor, or the University~~

- ~~Publication would preclude future publication~~
  - ~~Publication would be in breach of laws or ethics~~
- c) ~~Permanent or longer term embargo on all or part of print copy for a period of ... years (the request will be referred to the Pro-Provost and permission will be granted only in exceptional circumstances).~~

**Supporting statement for printed embargo request:**

ELECTRONIC COPY

- a) No embargo on electronic copy
- b) ~~Embargo on all or part of electronic copy for a period of ... years (maximum five) on the following ground(s):~~
- ~~Publication would be commercially damaging to the researcher, or to the supervisor, or the University~~
  - ~~Publication would preclude future publication~~
  - ~~Publication would be in breach of law or ethics~~
- c) ~~Permanent or longer term embargo on all or part of electronic copy for a period of ... years (the request will be referred to the Pro-Provost and permission will be granted only in exceptional circumstances).~~

**Supporting statement for electronic embargo request:**

Date ..... signature of candidate .....  
supervisor .....

signature of

*Please note initial embargos can be requested for a maximum of five years. An embargo on a thesis submitted to the Faculty of Science or Medicine is rarely granted for more than two years in the first instance, without good justification. The Library will not lift an embargo before confirming with the student and supervisor that they do not intend to request a continuation. In the absence of an agreed response from both student and supervisor, the Head of School will be consulted. Please note that the total period of an embargo, including any continuation, is not expected to exceed ten years.*

*Where part of a thesis is to be embargoed, please specify the part and the reason.*

## Abstract:

This dissertation is primarily concerned with the subjects of disagreement, argument, and the methodology of philosophy. The first chapter sets out and attempts to answer the question of what the connection between disagreement and disputing is. The second chapter is primarily a investigation into the nature of verbal disputes. The answer the chapter puts forward is that there is a justificatory relation (or at least we behave as if there is one) between disagreeing and disputing, so that, for example, if two parties do not disagree in the right way, then they (*prima facie*) should not dispute. In the second chapter I will look at a few theories of verbal disputes, and I will discuss some of the features such a theory should have. I go on to explicitly endorse a version of David Chalmers's theory of verbal disputes, and defend it from some potential objections. The third chapter is a defence of the method of conceptual analysis in philosophy. I introduce some potential objections to the Canberra plan style of conceptual analysis, and show how a different conception of conceptual analyses could get over these problems. The conception of conceptual analysis I argue for is heavily inspired by Rudolf Carnap's system of *explication*. The main way Carnapian explication would differ from the Canberra plan style of conceptual analysis is in the way that it would allow one to move further away from the original concept in analysing it, by balancing closeness to the original concept against other specific criteria.

## Acknowledgements:

Thank you very much to my supervisors who gave me excellent feedback on many drafts. I would particularly like to thank my primary supervisor Jessica. I would also like to thank my secondary supervisor Brian, and my previous supervisor for the MLitt dissertation Ephraim.

# Disagreement and Philosophical Method

## Introduction

This dissertation is primarily concerned with the subjects of disagreement, argument, and the methodology of philosophy. The first chapter sets out and attempts to answer the question of what the connection between disagreement and disputing is. In other words, it deals with the relation between having incompatible mental states and explicit argument. The answer the chapter puts forward is that there is a justificatory relation (or at least we behave as if there is one) between disagreeing and disputing, so that, for example, if two parties do not disagree in the right way, then they (*prima facie*) should not dispute. The explanation of this is the conventional acceptance of a norm rather like *resolution*:

*Resolution*: The default purpose of disputing is the resolution of disagreement.

The extent to which we see a dispute as justified would to some extent reflect the extent to which it was appropriate for resolving disagreement.

The second chapter is primarily a investigation into the nature of verbal disputes. In that chapter I will look at a few theories of verbal disputes, and I will discuss some of the features such a theory should have. I go on to explicitly endorse a version of David Chalmers's theory of verbal disputes, and defend it from some potential objections. My defence of Chalmers's account centres on his characterisation of verbal disputes as arising 'in virtue of' metalinguistic

disagreement.

The third chapter is a defence of the method of conceptual analysis in philosophy. I introduce some potential objections to the Canberra plan style of conceptual analysis, and show how a different conception of conceptual analyses could get over these problems. The conception of conceptual analysis I argue for is heavily inspired by Rudolf Carnap's system of *explication*. The main way Carnapian explication would differ from the Canberra plan style of conceptual analysis is in the way that it would allow one to move further away from the original concept in analysing it.

These chapters are by and large stand-alone works. However, there is a common theme running through all three. This is the Chalmers-inspired idea that it is where a dispute arises from that determines whether it is substantive or not. A first order dispute may be unjustified because it arises from second order disagreement. For example, a first order dispute over whether there was Champagne served at the wedding may arise from metalinguistic disagreement over what the word 'Champagne' means, where one disputant takes Prosecco to count as Champagne and the other doesn't. In the first chapter I argue—sometimes using intuitions about verbal disputes—that a dispute, in order to be justified, must arise from disagreement relevant to it, and I explain this by appealing to our acceptance of the *resolution* norm. In the second chapter I use these ideas from the first chapter to lend support to Chalmers's account of verbal disputes. In the third chapter, one of the objections I address is that conceptual analysis is a verbal enterprise, so that the choice between two competing analyses of a concept would be a merely verbal one. Part of the way in which I address this objection is to appeal to the earlier idea that it is where a particular dispute arises from that determines where it is verbal. I argue that if two conceptual analysts who dispute over whose analysis is better, are self conscious about the conceptual nature of their disagreement, then their dispute may be very much like a verbal dispute, but not viciously so.

## 1

# How disagreements ground disputes

Writers have realised there is some connection between disagreeing and disputing. Stevenson states in his account of disagreement that the parties will not be content to let the attitude they disagree with go unchanged or unchallenged (1944: 3). Allan Gibbard (2003) argues that holding incompatible plans is tantamount to disagreement because it can lead to normative discussion (at least a large part of this discussion should be considered disputing, broadly construed, for the people involved are putting forward seemingly incompatible hypotheses and arguing for them). Mike Ridge attempts to preserve the Gibbard's idea of understanding disagreement in terms of normative discussion in his own account (2013: 54). Huw Price states that disagreement is the “grounds for a dispute” (1988: 146). I take it that—implicitly at least—we as occasional disputers recognise that a dispute without the right kind of disagreement is unjustified, and also that if a dispute arises from some relevant disagreement then generally it *is* justified. This should not be controversial. When we notice that two people are arguing but seem to agree on all relevant matters of fact—as in a verbal dispute—we are inclined to point out that their dispute is pointless, rather than substantive. We are not so inclined when it is clear that the disputants disagree over the subject of their argument, for that reason at least.

Although some relation between disagreement and disputes is acknowledged by the quoted writers, the precise connection between them has not thus far been spelled out in much detail. My intent in this chapter is to explain this connection, and to note some implications of the explanation I favour. I argue that the justificatory relation (or our impression of one) between

disagreeing and disputing comes from our conventional understanding of the purpose of disputing; that disputing is for the resolution of disagreement. Although we may argue with other specific purposes in mind—for instance, to influence, annoy, and impress people—, that disputing can fulfil these particular ends in the way that it does may depend on our acceptance of norms which entail that the default purpose of disputing is to resolve disagreement.

In this chapter I shall first characterise disagreement and disputes so as to give us an idea of the subject matter. I will then examine how one particularly influential account of disagreement interprets the relation between disagreeing and disputing. After this I will put forward an account on which disagreement is essential for understanding disputes, when it comes to disputes over matters of fact, and defend it from some potential objections. In the last part, I will apply the treatment to certain species of disagreement in pro-attitude.

## Disagreement and disputes: Some preliminaries

Before we go on to look at some accounts of disagreement, which take into account its relation to disputing, and before we provide a positive account of our own, we must clarify our target. In the next few sections I will distinguish disagreements from disputes, I will examine the types of mental states which can be involved in disagreement, and I will briefly discuss some properties of disagreement which should feature in any account of it.

### *Disagreements and disputes*

In everyday speech we often use the terms 'disagreement' and 'dispute' synonymously, so drawing a distinction between them may seem rather pedantic; in fact, I am sure the title of this chapter will seem like nonsense to many. But once we draw the distinction, the rationale will become clear.

Following a few other writers (Chalmers, 2011, Jenkins, 2014: 13), I will use the term 'dispute' for conflicts where we explicitly engage in some kind of argumentative activity. This can be verbal or it can be physical. Two people arguing over whether Australia is bigger than Canada will count as disputing, as will two people engaging in fisticuffs. It may be thought that the latter does not constitute a dispute, but The Oxford English Dictionary is on my side here. A physical fight is as much of a dispute as a verbal argument, though perhaps of a different kind. At any rate, I could just say what I am concerned when I talk about disputes is *explicit conflict* between multiple parties. Whether I am using the word 'dispute' doesn't matter very much at all for this purpose.

One paradigm form of dispute is where one person says a sentence and in reply another utters an apparently incompatible one. So one person could say "P" and another could say "Not-P". Bear in mind that I am not saying that it is necessary that the disputants assert contradictory propositions, as we see in the next chapter, there are disputants who do not do so, specifically verbal disputants. For a dispute it is enough that the parties appear to contradict one another. There are other forms of dispute, depending on what they are over; one person may say "I want P" another may say "Well I won't allow you to have P". Here, as in the last case, the defining feature is that the disputants appear to be contradicting one another; the difference is that they do not appear to be contradicting each other with putative statements of fact. They are instead contradicting each other in what states they wish to be realised. Of course, disputes can be friendly. Two people with different views may try and work together to a shared conclusion, but in the course of this, they will likely say things which seems to contradict what the other says.

On the other hand, 'Disagreement'—as I am using the word—refers to a relation that holds between agents as a result of their respective mental states. Essentially, the distinction between disputes and disagreements then is that

former is constituted by behaviour and the latter by the inner states of the people involved. Two parties need not be engaged in argument in order for them to disagree. Nor do two parties have to be in disagreement in order for them to be having a dispute.

Sometimes disputes and disagreements, as I have characterised them, are conflated. We occasionally speak of disagreement as an activity; it is not uncommon to hear phrases like “Stop disagreeing with me!” or “Why is it that you go along with what your friends say but you always disagree with me?” Now some philosophers (Cappelen and Hawthorne (2009: 60-1)) have noted this and also speak about a separate notion of disagreement as an activity, so that when one argues with another, they are in a sense disagreeing with them. I don't really see how disagreeing in this sense differs in any significant way from disputing, so I will just assume that such activities are disputes as I have characterised them.

### *The vehicles of disagreement*

In the previous section, we have already stated that disagreement between two people is constituted by their mental states, but there is the question of which *kinds* of mental state are eligible for constituting disagreement. As Michael Ridge points out (2013: 46), it is not merely a matter of any two states having incompatible contents. Two people could entertain contradictory hypotheses or have contradictory fantasies and yet not disagree. The most obvious mental states which we can disagree in are beliefs.

Generally, if one person believes P and another believes Not-P, then we will be inclined to say they disagree. However, a few writers have taken it that that we can disagree, not only in our beliefs, but also in our other mental states. C.L. Stevenson, for example, distinguishes *disagreement in belief* from *disagreement*

*in attitude*<sup>1</sup>. In the literature, the attitudes that Stevenson has in mind are sometimes called 'pro-attitudes'; when I speak of attitudes in this chapter, I will speak of attitudes broadly construed, that is, including beliefs, etc. unless I am explicitly speaking about disagreement in attitude, by which I mean, disagreement in *pro-attitude*. These pro-attitudes cover a fair few states of mind. They include hopes, desires, and preferences. It is hard to spell out exactly the difference between desiring X and having a preference for X, but it seems to me to be that the difference is more one of degree than of kind, so that a preference would perhaps be a weak desire. We can add further detail to our characterisation of pro-attitudes with the notion of *direction of fit*. I take pro-attitudes to be those attitudes with a world-to-mind—rather than a mind-to-world—direction of fit. The notion of direction of fit requires some explanation. Mark Platts says of it:

Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit the world, not *vice versa*. Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not a failing in the desire, and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not *vice versa*. (1979: 256-7).

A way of cashing out the precise difference between mental states with a world-to-mind direction of fit and those with a mind-to-world one, without using highly metaphorical language, is Michael Smith's account. He distinguishes a desire that P from a belief that P by the different counterfactuals true of them, regarding a holder of those attitudes perceiving that Not-P. If Someone who believes that P perceives that Not-P, then their belief that P will generally go out of existence. Whereas, when someone with a desire that P perceives that Not-P, their desire will normally endure (1992: 113-6). Perhaps this is not the defining feature of either beliefs or desires, but it shows that there are ways to distinguish between them.

---

1 Of course, accepting this distinction does not entail we must endorse Stevenson's particular account of either species of disagreement. In fact, in this chapter I accept that we can disagree in both beliefs and desires and then I go on to reject Stevenson's account of the mechanism by which we do.

That we can disagree in our full-blown desires is relatively uncontroversial. It is a view held by Frank Jackson (1985), Simon Blackburn (1998), Philip Pettit (1998), Torfinn Huvenes (2012, 2014), John Hawthorne (2014), and Michael Ridge (2014). Stevenson's examples of people disagreeing in their desires are also quite intuitive:

John's mother is concerned about the dangers of playing football, and doesn't want him to play. John, even though he agrees (in belief) about the dangers, wants to play anyhow. Again, they disagree. (1944: 3)

I personally find this quite compelling and would naturally call it a case of disagreement.

However, it is slightly more controversial that we can disagree in our preferences. Although some writers do hold this view (Weatherson (2009), Huvenes (2012, 2014)) it seems to some (including me) to be right on the boundary between disagreement and mere difference in mental state. I will not pursue the question of whether we can disagree in our preferences, or indeed, whether there can be cases of 'Faultless disagreement' in this chapter. For now we should agree on the non-controversial idea that we can disagree in both our beliefs and our desires.

### *Some features of disagreement*

In the last few sections we have already noted a few features of disagreement. For instance, that it is constituted by the mental states of the relevant parties rather than their behaviour, that it is constituted at least partly by the parties holding contrary attitudes, and that there are multiple types of mental state—e.g. beliefs, desires—of which tokens can constitute disagreement. There is one more feature of disagreement which we should mention, before going on to look at an actual account of disagreement and trying to explicate its connection to the notion of

disputing.

This feature of disagreement is its logical form. As John MacFarlane<sup>2</sup> points out

$x$  is in disagreement with  $y$

is not specific enough for an adequate account (2014: 120). For a start, X and Y need to be in disagreement *over something*; they won't be in disagreement simpliciter, whatever that might be. MacFarlane also holds that some kinds of disagreement lack propositional content, so it wouldn't be enough to say that two people disagree over a particular proposition P. One way to capture this is by holding that the relation is somehow disagreement with  $\Phi$ -ing; this  $\Phi$ -ing could be holding a particular belief or pro-attitude. Some disagreement, according to MacFarlane, also can depend not only “on the contents of the relevant attitudes, but on the contexts in which they occur” (ibid). A case of people holding attitudes with conflicting contents which *do not* constitute disagreement could be where Ambra at 3pm believes *Jane is sitting*, and Alessandro at 5pm believes *Jane is not sitting*. A case where we have disagreement but no conflicting contents could be where Ambra believes *Jane is in the hairdressers* when she is in St Andrews high street, and where Alessandro believes *Jane is in the hairdressers* when he is walking in Bologna. Disagreement is not merely a relation between multiple people; the logical form of disagreement is that of a “relation between a person and a possible speech act or attitude in context” (ibid). MacFarlane concludes that the logical form of disagreement is this:

$x$  is in disagreement with  $\Phi$ -ing-in-context- $c$

$\Phi$ -ing in this sense could be either performing a particular action, or it could be holding a particular belief. Despite this, I will mostly talk about two people disagreeing with each other, rather than each other's attitudes. I don't think we

---

2 In this section I use some controversial examples, specifically the ones which depend on thesis that Ambra and Alessandro could believe the same proposition (*Jane is in the hairdressers*) despite their different contexts. A contextualist, for example, could say that both believe different propositions, i.e. *Jane is in the hairdressers in St Andrews* and *Jane is in the hairdressers in Bologna*. I do not commit to any particular thesis on this subject; I use the example merely to illustrate how context is important to disagreement.

lose anything by speaking this way in most cases. However, we need to understand that disagreement is essentially a relation between agents and certain attitudes when we go on to talk about how part of disagreement is disapproving of the holding of that attitude. The relevance of this will become clear later, but for now it will suffice to say that 'disagreeing with  $\Phi$ -ing-in-context- $c$ ', better represents the way we are inclined to disapprove of the attitudes we do disagree with, than does the simple 'x is in disagreement with y' structure.

## Stevenson's account of disagreement

To really illustrate the question I wish to answer in this chapter, it will be useful to look at an account of disagreement which provides an answer (or rather, a sketch of one) to it. C.L. Stevenson provides such an account. Before I investigate how he links disagreement and disputing, I will outline his account. Stevenson attempts to capture both disagreement in belief and in attitude. Here are a few quotes of his which detail his theory:

'disagreement in belief' . . . occurs when Mr. A believes P, when Mr. B believes not-P, or something incompatible with P, and when neither is content to let the belief of the other remain unchallenged. (1963: 1)

[Disagreement in attitude] occurs when Mr. A has a favorable attitude to something, when Mr. B has an unfavorable attitude to it, and when neither is content to let the other's attitude remain unchanged (1963: 1)

The difference between these two senses of 'disagreement' is essentially this: the first involves an opposition of beliefs, both of which cannot be true, and the second involves an opposition of attitudes, both of which cannot be satisfied (1963: 2).

Stevenson's overall account of disagreement has two main components: First, incompatibility of the relevant attitudes (broadly construed to include beliefs).

Second, both parties wanting to change each other's attitudes.

His account distinguishes between disagreement in attitude and belief in a few ways: First, disagreement in attitude requires that one of the relevant parties has a favourable attitude to the thing they are disagreeing over whilst the other has an unfavourable attitude to it; so one party may have a desire for something and another would have a desire against that. This would not be the case with disagreement in belief; a belief should not be construed as a favourable attitude to something in the same way. A second related difference is the way the two types of disagreement feature incompatibility. Disagreement in belief features beliefs which cannot both be true, whilst disagreement in attitude features attitudes which cannot both be satisfied. The difference here is in the mental state's direction of fit; the respective type of mental state's teleology.

### *Stevenson on disagreement and disputes*

Recall that earlier on in the chapter (in the section entitled “Disagreements and disputes”) we noted the importance of separating disagreements and disputes so that we do not conflate them. But in another sense, it is important to link them, which is the subject of this chapter. The former sense is that disputing and disagreeing are different notions, disputing being overtly behavioural and disagreeing being primarily a relation between mental states. The latter sense concerns the degree to which the notions themselves are related; saying that there are norms entailing that disputing is only justified when there is disagreement present would be linking the notions in this sense.

Stevenson's account is primarily of interest to us because of the way his account links disputing with disagreeing. The way he does this is summed up by two clauses contained respectively within the first and second quotes. The first is that when two people disagree in belief they are not content to allow each other's beliefs to go unchallenged, the second is that when two people disagree in attitude they are not content to let each other's desires remain unchanged. These

both have the implication that when they disagree, the parties will view one another as having the wrong attitudes and being in need of correction; at the very least, the parties would view each other's attitudes as worthy of being called into question. This is where disputing comes in; the natural way of changing or challenging another person's attitudes is to argue with them. With this method we can criticise the position of the person with whom we disagree; if they believe that P, and we do not, we could present evidence against P holding to change their mind. This is most apparent in cases of disputes over seeming matters of fact. One person says "P", another person believes that Not-P and because they take the first person to be incorrect they challenge them by saying something to the effect of "Not-P". This is the way Stevenson links up disagreements with disputes; Disagreement brings with it a *prima facie* desire to dispute with whoever we disagree with.

It might be thought that Stevenson is just conflating disagreeing and disputing; so that in order to disagree with someone, we have to also dispute with them. This would be false however. Notice that according to the first two quotes, one does not need explicit conflict in order to disagree with another; rather, one needs a motive for arguing with them—e.g. To change or challenge the attitude they hold, because they are discontented with them holding it. In this way, his account is perfectly consistent with the notions of disagreement and dispute being distinct. One can satisfy Stevenson's conditions for disagreement without exhibiting argumentative behaviour. In this way, not only does Stevenson's account allow the notions of disagreement and dispute to be distinct, it also shows how they are related.

However, disagreement, as I have said, is a relation that can hold between two people regardless of familiarity or spatio-temporal location. Recognising that someone holds a particular attitude and wanting to change or challenge it is not necessary for disagreeing with them. If it was, then it would be wrong to say a perfectly respectable sentence like "I disagree with Mussolini on a great many things, but..." or "I disagree with anyone who believes in ghosts!" The fact that these sentences don't strike us as infelicitous, and that we often do use the notion of disagreement in this way is good evidence for the claim that disagreement does

not require familiarity. What Stevenson's account gets wrong is it entails that in order to disagree with someone we need to be uncomfortable with that person *actually holding those attitudes* we disagree with. Recall John MacFarlane's view that disagreement is a "relation between a person and a possible speech act or attitude in context" (2014: 120). This entails that we can disagree, not only with people we are unfamiliar with, but also with certain attitudes. For example, we can make statements such as "I completely disagree with the view that Tony Blair is secretly from Kazakhstan" and not be saying nonsense. In this way, if someone holds a particular attitude in a certain context, then we can disagree with them without knowing they hold it. Hence dissatisfaction with letting a particular person's attitudes remain unchallenged or unchanged should not be a necessary condition for disagreement.

Stevenson's conditions get something important right though. Generally, disagreement is uncomfortable; it implies that someone is at fault. This is especially apparent when two people disagree in belief. In many such cases, at least one person will be not merely false, but also *wrong*. To jump ahead slightly, the presence of disagreement brings with it an incentive for it to be resolved. This could be the reason for us generally viewing disagreement as a matter to be resolved, preferably in favour of our own point of view. But as we have seen, what Stevenson's condition gets wrong is that the relevant disapproval is not limited to cases where the parties believe of each other that they hold incompatible attitudes. An agent can disagree with someone without being at all familiar with them; they can even disagree with attitudes that no one holds. To go back to MacFarlane, this is because disagreement is a relation between an agent and a possible attitude or speech act. When we disagree with someone, we disagree with the attitude they hold.

Part of disagreeing with an attitude would be disapproving of someone holding it. This is why we accept a slightly weaker claim that when two people *know* they disagree they will not be content to let each other's attitudes go unchallenged; each person would disapprove of the relevant attitude the other holds. However, Stevenson is wrong in holding the stronger position that in order to disagree, two people must want to actually change the other's attitudes. The

truth of the weaker claim could just be the consequence of agents disapproving of the holding of attitudes when they disagree with them, and hence viewing others holding these attitudes as being in need of correction. In this sense, us not being content to let people remain unchallenged when we know we disagree with them, could be a case of a more general phenomenon.

So what Stevenson gets right is that the we are not content to live and let live when we know we disagree with someone. What he gets wrong however, is that disagreement stretches further than those cases where we actually want to change someone's attitude. In the next few sections I will attempt to provide a different and more detailed account, which is consistent with disagreement being able to hold between agents who are unfamiliar with each other.

## Grounding disputes

Now I wish to examine the connection between disputes and disagreement in more detail. As stated in the introduction, I believe the connection between disagreements and disputes is justificatory in nature, so that (roughly) a dispute is justified when it arises from the appropriate type of disagreement, and unjustified when it does not. At the very least, our practice of disputing suggests strongly that we view disputes not standing in the appropriate relation to disagreement as unjustified. For instance, take this verbal dispute: A British person—referring to a crane fly—could say to an American “There was a daddy-long-legs in the kitchen, but it flew out of the window”, and the American—believing that the Briton is referring to a harvestman—may reply “No it didn't, daddy-long-legs can't fly!” We would not view this dispute as justified, as it arises from them using 'daddy-long-legs' in different ways, rather than disagreement over whether a crane fly (or a harvestman) flew out of the kitchen window. One may characterise the Briton and American as having metalinguistic disagreement, that is, conflicting beliefs about what the phrase 'daddy-long-legs' refers to. However this

disagreement would not be of the appropriate type to justify the dispute.

The appropriate type of disagreement I take to be disagreement which is relevant to the subject matter of the dispute. The Briton and the American may have metalinguistic disagreement over what 'daddy-long-legs' refers to, and this disagreement could surely ground a dispute over this subject; however, their dispute is a first order one about whether a certain type of animal can fly, and would have to arise from disagreement over this matter in order to be justified. As it happens, it doesn't arise from this source—rather it comes from their linguistic differences, which would give rise to the dispute whether or not the disputants believed all the same things about the respective biologies of crane flies and harvestmen.

This justificatory relation I envision is slightly different from that proposed by Stevenson's account; the latter implying that we are motivated to dispute when we disagree, but does not say much more on the subject than that. The justificatory account says instead that two people have a (at least *prima facie*) reason to dispute when they disagree, and that they have no reason to if they don't; if two people do not disagree, then they should not dispute. The difference between the the justificatory relation and Stevenson's is that in order to meet Stevenson's condition (that we are motivated to change the attitudes of the person we disagree with) one must be familiar with the person one disagrees with, so that one knows that together they hold jointly contrary attitudes; whereas one can have justification to dispute with another regardless of knowledge of them or their mental states. The justificatory condition I envisage, in contrast, could hold regardless of one's knowledge of the mental state of who they disagree with. For example, if I disagree with someone whom I am not familiar with over whether David Cameron will win the next election, I have a reason to dispute with them. Of course, there may be other overriding reasons for me not to dispute with them, e.g. that I do not know that they disagree with me, that we are in a no-talking zone, that we don't actually know each other, etc., but these do not mean there is no *prima facie* justification for arguing, which can be overridden by others.

My explanation of this apparent justificatory relation, is that there is a certain default purpose that disputing (at least conventionally) serves for us, and

the extent to which we view the existence of a given dispute as justified is tied to the way in which the dispute is suited to serving this purpose. This purpose of disputing, I argue, is to resolve disagreement. It is for this reason that we see disputes which do not arise from appropriate disagreement as *prima facie* unjustified, and disagreements that do as *prima facie* justified. What I mean by 'default purpose' here does not entail a metaphysical thesis about the actual and possible structure of disputes. My point is just that there is a conventional understanding that if two people are sincerely disputing, then generally they will be working towards agreement. As I shall show, there are cases where disputes may not be aimed at this goal, but most—if not all—will feature insincerity on the part of the person with this other goal in mind; one disputant will try to give the impression that there is disagreement that they are trying to resolve. Although we can—and do—dispute for other purposes, resolving disagreement is viewed by us as the primary purpose of disputing, and perhaps other possible purposes it may serve are derivative of this. We behave as if there is a certain norm governing our argumentative practices. *Resolution* I think captures this:

*Resolution*: The default purpose of disputing is taken to be the resolution of disagreement.

Now in supporting the claim that our disputing behaviour is in accordance with *Resolution*, I will restrict my attention to disputes over apparent matters of fact; those seemingly grounded by conflicting beliefs. That we have genuine disagreements in the vicinity is uncontroversial (unlike when talking about 'disagreement in preference'). I will wait until later to apply the treatment to other types of dispute and disagreement. I shall first explain *Resolution* and show how we do behave as if we are governed by such a norm. Later on, I will show how this can support the idea that the relation between disagreeing and disputing is a justificatory one.

I will occasionally slip between talking about us *seeing* disputes as justified and them *being* justified. This is mainly for sake of ease. When I do say that a dispute is justified, I mean that it is of the type we are liable to see as justified. One may object that, in that case, all I am showing in this chapter is not anything interesting about disputes, but rather just some observations about our disputing

behaviour. The interesting question, they may add, regards how to tell whether a dispute *really is* justified. I however, think there are good reasons to be interested in the way our folk-notions of disputing and disagreeing interplay. As I state in the next sections, there are good reasons for resolving disagreement, and disputing is the best way for us to do this. The way we see disputes as justified, may heavily imply that they are justified, as the conventional purpose they hold for us clearly at least has some pragmatic virtue.

### *The default purpose of disputing*

So what support do we have for the claim that the default purpose of disputing is to resolve disagreement? Well let's first start from the uncontroversial point that we frequently do argue with this purpose in mind. When two parties argue about a matter of fact, such as whether or not King Arthur was a living person, it is a pretty safe bet that they are trying to bring the other person around to their point of view, and hence establish agreement on it. But even if the parties are not *consciously* trying to establish it, agreement can still be the product of reasoned argument; also resolving disagreement in this way has many benefits. Huw Price produces some valuable insights into the relation between disputes and disagreement in *Facts and the Function of Truth*. Price argues that disputes are behaviourally advantageous because they allow us to assess a range of options of how to act, or rather assess beliefs with different behavioural consequences. When we are in a community we have access to a greater body of experience than we would if solitary. It can be useful for us to have our beliefs tested when we have access to the community's experience; a better informed person can tell us that the attitudes we hold may lead us to perform actions with bad consequences, e.g. injury or death. A dispute consisting of reasoned argument seems to be the natural way to decide who is correct in situations where our attitudes seemingly conflict, as *ceteris paribus* the person who is correct (or perhaps

just better informed) will produce the more convincing case. Price says:

In the long run, I suggested, such disputes can be expected to have a beneficial effect on the behavioural dispositions of individual speakers. They help us ensure that as individuals we hold and act on attitudes that reflect, to some extent, the combined wisdom of our linguistic community. Our behavioural dispositions can thus be tested against those of other speakers, before they are put to use in the world. The guiding principle is that it is better to be criticised for claiming that tigers are harmless than to discover one's mistake in the flesh (1988: 145).

We can see from this quote that Price holds that certain disputes have a practical basis, in that they encourage good behaviour and discourage bad behaviour. In a dispute we would investigate which disputant has the more robust view by seeing which stands up to the most scrutiny. Agreement would be established by one disputant relinquishing their view and adopting the better one. Agreement is not good for its own sake, for we can agree on falsities; it would not do very well for the disputants to argue over whether Tigers are harmless and then conclude that they indeed are. In disputing we should want to establish good behaviour all round, and hence we would want to establish agreement on what is correct to believe.

Resolving disagreement via disputing can have this result because in a speech community, “utterances of a given sentence tend to be correlated to mental states with similar causal or functional roles in the determination of behaviour” (1988: 152). Price calls the *Same Boat Property* the one which is shared by classes of mental states whose “behavioural appropriateness, or utility, is predominantly similar across a speech community. If a mental state has [this property], then if it is appropriate for any one of us, it is appropriate for all of us” (ibid). Presumably the beliefs we agree on in certain disputes have this property, which is a reason for wanting agreement; it would establish good behaviour all round. Furthermore, this would mean that when two people are disputing—as it is likely they really do disagree—they would be in different mental states and so one would likely be behaviourally disadvantaged “If a mental state has the SBP, then if it is appropriate for any one of us, it is appropriate for all – we are all in the same boat” (ibid).

One explanation for this is that the useful beliefs—the ones which dispose us towards good behaviour—are generally true, or have some truth-seeking property. Good behaviour will then be that which arises from having true beliefs; the ones which will help us in our goals, e.g. surviving and succeeding. This could also be thought to be what underlies the same boat property: “it may seem that the required case of the SBP is a trivial consequence of the truth-conditional view of belief; that beliefs with the same truth condition are clearly appropriate in the same circumstances – namely, when and only when their truth conditions obtain” (ibid). So in aiming at establishing good behaviour all round, disputes aim at everyone getting to the truth of the matter; it cannot do that unless it resolves the disagreement. Now for our purposes we could perhaps go in for this idea, but Price would view this explanation as putting the cart before the horse. His explanation of truth is centred around its function in rational argument, and furthermore, is a reaction to what he sees as a failure to provide an analysis of truth. For now, I will not explain the usefulness of disputes in terms of truth, mainly because I don't think it is necessary once we accept that the appropriate behaviour will be the kind which arises from beliefs which are more likely to stand up to scrutiny than those which produce inappropriate behaviour.

Our motivation to establish agreement on the attitudes which will give rise to good behaviour, Price says, comes from our acceptance of what he calls *The Truth Norm*:

Truth: If Not-P, then it is *incorrect* to assert that P; if Not P, there are prima facie grounds for censure of an assertion that P. (2003: 170).

This norm entails that if we believe that Not-P and we hear someone assert that P, then we take it that that person is at fault, and hence, we are disposed to disapprove of them believing Not-P. This, Price states, gives us “preferential pressure towards settling the disagreement in question” (2003: 175). This norm, he says:

makes what would otherwise be no-fault disagreements into unstable social situations, whose instability is resolved only by argument and consequent agreement—and it provides an immediate incentive for argument, in that it holds out to the successful arguer the reward consisting in her community's positive

evaluation of her dialectical position. (ibid).

Of course, being rewarded by one's community is not the only motivation for settling disagreement. Our disapproval itself will make us want to explicitly dissent from opinions we do not agree with; in general, disagreement just does not sit comfortably with us.

Now I think that Price is right, when it comes to the utility of disputes I have restricted the treatment to for now—that is, those over apparent matters of fact. But these remarks by themselves only establish that disputing is a useful exercise as it can resolve uncomfortable social situations and encourage useful behaviour. In order to show that the default purpose of disputing is to resolve disagreement we need further argument. Here are two points that I believe support this conclusion: Firstly, when two disputants find out that they do not disagree, they will most likely finish arguing about whatever it was they thought they disagreed over—if they really want to argue for whatever reason, they may shift the focus of the dispute rather than finish arguing altogether. We can see this in further examples of verbal disputes (the subject of the next chapter). For instance, if a British person says to their American partner that they “need new pants” when they have very little underwear in their wardrobe but an abundance of trousers, the latter may dispute what the former says by replying “But you have tons of pants”. However, when both are shown that they each use the term 'pants' in a different manner and that they agree on the respective numbers of underwear and trousers in the wardrobe, they would generally cease arguing about the subject at hand. They may carry on arguing, but about a different subject; they may follow up their original dispute with one about what the word 'pants' denotes in the linguistic community they inhabit, or even about whether the number of underwear in the wardrobe is sufficient. This does not entail that we *always* dispute to resolve disagreement, but it does suggest that we accept certain norms about when to dispute. For instance, one such norm—which we have previously alluded to when talking about the justificatory nature of the relation between disagreeing and disputing—is this:

*Disagreement 1*: Two parties should (prima facie) not dispute with another if they do not disagree over the subject matter of the dispute.

The best explanation of such norms would be us accepting something like *Resolution*. It is all very well *Truth* leaving room for resolving disagreement and for motivating it, but we would need another norm to place the presence of disagreement as a restriction on disputing. Now, there is a variation of *Disagreement* we would not accept, i.e. that if the two parties do disagree then they *should* dispute. As I stated earlier, there may be other overriding reasons not to dispute. What we would accept, is this:

*Disagreement2*: If two parties disagree, then they have a reason to dispute with one another.

So if two people disagree, then we have the grounds for a dispute. *Truth* pretty much gives us this result: if two people genuinely disagree (recall, for the moment we are speaking only about disagreement over apparent matters of fact; I intend to exclude disagreement in attitude, and faultless disagreement) then one of them will be incorrect, and so there will be “prima facie grounds for a censure” of the incorrect view (2003: 170); we will want to resolve the disagreement in favour of the correct view.

The second point I will make can be seen as an answer to a potential objection as well as a positive argument for my position. The objection would go something like this:

You say that the default purpose of disputing is to resolve disagreement. Well there are plenty of ways we can dispute with a different purpose in mind. I may argue with someone because it will annoy them, to impress an onlooker, to put them in their place. There are situations where I may play devil's advocate, or argue with someone who I know will never agree with me. How then can we say that resolving disagreement is the default purpose of disputing?

My answer to this potential objection is that although particular disputes may not themselves be specifically aimed at resolving disagreement, they fulfil their intended purposes because of our mutual understanding that the purpose of disputes is to resolve disagreement in favour of the correct view. For instance, one of the reasons disputing with someone (under certain circumstances) may annoy them is because it heavily implies that we take them to believe or be saying

something false; falsity is a defect when it comes to assertions and beliefs, and they may be annoyed that we would have the audacity to question their beliefs, or that they may be wrong.

By the same token, we could impress an onlooker by arguing with someone else because in doing so (and doing so well) we could give the impression that we are right and our fellow disputant is wrong. This would give suggest that we are in a better epistemic position than our opponent. We could put someone in their place by disputing with them because in doing so we may give the impression that our beliefs are correct whilst theirs are false (again, if we argue well). The intended effect would be maximised in these cases if we won the argument, and hence established agreement in favour of our view.

So in some cases, the conventional purpose of disputing being to resolve disagreement is compatible with us occasionally disputing with a different purpose in mind. Furthermore, as we see in the previous few paragraphs to some extent disputing to achieve these purposes may rely on this mutual understanding of the purpose of disputing. It is for these reasons I do not think pointing to particular intended effects of a dispute is a good strategy to challenge *Resolution*. The problem with this strategy is that *Resolution* does not entail that we *only* dispute with the intended purpose of resolving disagreement; it is rather a claim about what we conventionally recognise the purpose of disputing to be.

A better strategy would be to produce a convincing case where someone argues in a way which relies on the conventional understanding of the purpose of disputes being something other than establishing agreement. To me it does not seem likely that disputes over matters of fact could conventionally be used for any purpose significantly far away from this. There is something about the structure of disputes over factual matters, which makes it fit for resolving disagreement in favour of the correct view. I believe P, you believe Not-P; I produce evidence for P, you do the same for Not-P. The presentation of our respective positions, our attempts to justify them, and criticism of the other's view, makes it seem very much like we are trying to increase credence in our position and lower it in others. If this credence lowering and raising is carried out to a certain point, then the disagreement is resolved. Generally, but not always, better arguments and

more justified positions will be more compelling than the weaker arguments and the less justified positions. So we can see how this process of disputing can lead to agreement on the best position, and how it seems like a rather good way of getting to this point.

Even if we did generally use disputes over matters of fact to gain influence, win favour, etc., could we understand disputes to be for this purpose? Such a situation would beg the question of why others would place value on success in a dispute at all. We have a clear rationale for it if *Resolution* is correct, as we would understand success in disputes to be generally indicative of correct beliefs. But if we viewed the purpose of disputing as simply gaining favour, then we would need to ask what we would be winning the favour for.

In the case of playing devil's advocate. In doing so, we would assume that if one position (the devil's position) were correct, then the other (the angel's) would be false. One would make as if there is disagreement between the two disputants and see where the argument between them would go. One must remember that playing devils advocate is not an end in itself, rather it is itself a conversational tool. So in order to see how it can rely on a recognition of what the default purpose of disputing is, one must see what we try to achieve by playing devils advocate. Now one might—just as in the first case we looked at—be trying to agitate someone by denying what they say, but we have seen already how cases like this fit into my account. What other value does it have? Well it means that in a room full of people who agree on P, we can still get the benefits of argument, e.g. it can force us to justify our own opinions and it can help us consider other points of view. It is a way of simulating a dispute grounded in real disagreement, so that we can get some of the benefits that argumentation brings. In cases like this, we would still get the best result from establishing agreement on the better point of view; part of the benefit of simulating disagreement, is that you act as if there is some present, and you aim to resolve it. It is in this way that playing Devil's advocate relies on the default purpose of disputing to be resolving disagreement. Playing out an argument between the two viewpoints, as if it was a genuine dispute, gets us the most from the exercise.

The same goes for disputing with someone who will not change their

mind. We must ask for the reason one has in arguing with someone so set in their own views. True, if we know that someone wouldn't change their mind even if we did argue with them, then it would be irrational to attempt to bring them around to our view. But it is not incompatible with *Resolution* to attempt to do so, especially so when we are inclined to disapprove of those who we disagree with; someone may still argue with another who they know won't change their mind, because they view attempting to correct someone as the right thing to do, even if they are unsuccessful at it. In order to show such a situation to be a counterexample to our treatment of disputing, our interlocutor would have to provide a plausible purpose for doing so which was incompatible with *Resolution*. For the reasons given in this section, I am doubtful that one can be found.

### *Consequences of this*

There are some novel, but I think intuitive, consequences of *Resolution*, some we have already mentioned: (1) It gives rise to certain derivative norms governing our practices of disputing. For instance, it strongly suggests the general acceptance of norms like *Disagreement*, which entail that in order to be prima facie justified, a dispute must be grounded in disagreement. But *Resolution* I think suggests a stronger conclusion than this. It suggests my earlier assertion (2) that not just any old disagreement can justify a dispute. Our disagreement over whether the poem *The Battle of Maldon* is a reliable historical source, would in no way justify a dispute over whether the X60 or the 95 bus is the faster way to travel from St Andrews to Leven. However, it may justify a dispute over whether the battle really took place next to the Blackwater river. This is obvious, but there is a reason for this which draws from *Resolution*; the dispute over the buses could not, or should not, help resolve any disagreement over the battle of Maldon. This is because of the two subjects' irrelevance to one another.

So the disagreement must be *relevant* to the dispute. I would go further, and

say that the disagreement must somehow *explain* the dispute. It wouldn't do very well for there to be a dispute between two people where they disagree over the relevant matters but where the disagreement in no way causes or contributes to the dispute; the dispute must arise, or at least progress, in virtue of the underlying disagreement. Observe David Chalmers's definition of a verbal dispute:

A dispute over S is broadly verbal when, for some expression T in S, the parties disagree about the meaning of T, and the dispute over S arises wholly in virtue of this disagreement regarding T (2011: 522).

He clarifies the in-virtue-of characterisation by saying it “should be understood as an explanatory 'in virtue of': the idea is that the metalinguistic disagreement explains the apparent first-order disagreement. And the relevant sort of explanation should require something stronger than an arbitrary causal or evidential explanation” (2011: 523). Just as a verbal dispute would arise in virtue of metalinguistic disagreement<sup>3</sup>, a justified dispute would be explained by the disagreement between the disputants.

## Disagreement in attitude

In this section I will show how it is possible to extend my account to certain cases of disagreement in pro-attitudes. Thus far I have explicitly restricted my account to disputes and disagreements over apparent matters of fact. However, as we see from early on in this chapter, this is not comprehensive of disagreement in general. In a dispute over matters of fact we call into question the other disputant's *beliefs*; accordingly disagreements which ground such disputes will be cases of disagreement in belief. But as we have seen from Stevenson, we can disagree *in attitude*—or rather, *pro-attitude*—as well. As our account stands, we haven't answered the question of how disagreement in attitude can ground

---

<sup>3</sup> Something I call into question in the second chapter.

disputes. Some norms such as *Truth* that I mention in the previous sections do not seem to have much crossover at all to disagreements between attitudes like desires; desires just aren't the kind of things that are capable of being true or false.

Saying this, I think our account can encompass other types of disagreement with some tweaking. The main thrust of the last few sections was that the link between disagreeing and disputing is a justificatory one. We can still get that same result in cases of disagreement in attitude, although what makes disagreements in attitude justify certain disputes will differ from disagreement in belief. Just as in the cases of disagreement in belief, when disagreement in attitude—or rather, the cases I will limit myself to—underlies a dispute, it should be appropriate for the purpose of grounding that particular dispute; it should be *relevant* and should *explain* the dispute. The way the justificatory relation holds between certain cases of disagreement in attitude would reflect the purpose—or conventional purpose—of the disputes that they ground; this purpose however, will perhaps differ from those grounded by disagreements in belief.

These points form the core of my explanation of the link between disagreeing and disputing. The justificatory account, I shall show, can be applied to at least some disagreements in pro-attitudes. For some types of disagreement in attitude, the disagreement can justify certain types of disputes, and those types of disputes should not be undertaken if not grounded by the right kind of disagreement in attitude. The brief treatment I will give in this section is not meant to cover *all* disagreements in attitude. Rather, its purpose is to show that something like the treatment of disagreements and disputes over apparently factual matters that I gave in the previous sections can, in principle, be extended to other types of disagreement and disputes. Our treatment of disagreement in attitude will be slightly different from the former, but not in a way that jeopardises the broad overall account.

Let's consider an example of disagreement in attitude from Stevenson:

John's mother is concerned about the dangers of playing football, and doesn't want him to play. John, even though he agrees (in belief) about the dangers, wants to play anyhow. (1944: 2)

Now, a dispute could very easily arise in virtue of this case of disagreement in

attitude, but this dispute would differ in a number of ways from a dispute over matters of fact. Firstly, it would differ in the way the dispute could be won. In a dispute over matters of fact, one party wins the dispute when they bring the other around to their view; it requires a change in belief on behalf of one disputant. However, in the football case, the dispute can be seemingly be won without a change in the relevant attitude. Perhaps John's mother, while not changing John's attitude about playing football, still manages to persuade him not to play. She could change his beliefs regarding how much it would upset her if he injured himself, or she could even attempt to increase the potency of his desire to not upset her without changing his beliefs. The aim of both of these options here would be to override John's desire to play football with another desire, in particular, to not upset his mother.

It is clear to see that there is a similar justificatory link here, as in the cases discussed in the last few sections. If John and his Mother did not disagree in attitude, but for some reason still engaged in the same dispute (perhaps with John lying about wanting to play football to annoy his mother), then we would be inclined to view this dispute as ungrounded. So, at a first glance, it may seem as if our practice of disputing in cases like this, is governed by a norm such as *Resolution*. So that we see the John and his mother's original dispute as justified because of their dispute being appropriate for the purpose of resolving disagreement in attitude; in the cases where they do not disagree in attitude, we would not see the dispute as justified because of there being no disagreement to resolve.

The first dispute (where John and his mother disagree in attitude) seems justified because of the presence of disagreement, and the second (where John and his mother do not disagree in attitude) seems unjustified because of the lack of it. However, this may not entail that these kinds of disputes are geared towards resolving disagreement. Sure, if agreement is established in favour of one view then one party has been successful. But this may not be the only way to be successful in a dispute. Consider success in the kinds of disputes we looked at in the previous sections, those grounded by disagreements in belief. Our beliefs are supposed to capture how the world actually is because they have a mind-to-world

direction of fit. In this way, when two parties are arguing over what is the case, they are working out which beliefs fit the world (or perhaps which ones are the most useful for navigating it) and therefore, which ones to have. Success in an argument is it ending in favour of what one takes to be the correct beliefs.

But the case of John and his mother is different diverges from this template because their dispute is grounded by incompatible desires rather than beliefs. Now, success in a dispute about apparent matters of facts requires changing one's opponent's beliefs; the dispute is about whose beliefs fit the world, and hence whose beliefs are correct. However, desires have a world-to-mind direction of fit, rather than a mind-to-world direction of fit like beliefs. The way in which desires are incompatible—at least in the case we are looking at—is that they cannot be jointly satisfied<sup>4</sup>, whereas beliefs are incompatible when they both cannot be true. Another way of putting this is that two beliefs are compatible when they both cannot fit the world, but two desires are incompatible when the world cannot fit both of them. This suggests that success in a dispute grounded by incompatible desires, rather than beliefs, requires deciding whose desires the world will fit; which state of affairs *is to be* realised rather than what state of affairs *is* realised.

As I see things, this does not entail one disputant giving up their competing desire; the disputants could decide which world state is to be brought about, without one changing their mind and actually wanting that state of affairs. Two people could argue over what album to listen to in the car, but then decide that the driver should get to choose. In this case the passenger may still despise the driver's music and much prefer the situation where the don't have to listen to it. The driver has still won the argument. Similarly, in the case of John and his mother, John does not need to relinquish his desire to play football in order for his mother to be successful in the argument; he just needs to not play football. So in cases grounded by jointly-incompatible desires, it seems that success in the argument comes not from the two disputants changing their attitudes so they

---

<sup>4</sup> This is not the only way that desires can be incompatible. What I am advocating here is something close to Stevenson's conception of disagreement in attitude.

align with each other's, but rather, establishing whose desires get satisfied<sup>5</sup>. In this way, the aim of the dispute, because of the direction of fit of the attitudes that ground it, may be to decide on which state of affairs is to be realised.

There may still be preferential pressure to change the other disputant's attitudes, but this may come from the fact that changing a person's attitudes will lead to the state of affairs we prefer being realised. Desires themselves provide a motivation to bring about certain situations, and part of bringing about your preferred situation may be stopping other people from bringing about incompatible situations. Perhaps this is generalised into a particular norm, for example:

If two people disagree in attitude, then they each have a *prima facie* reason to change the other person's relevant attitudes<sup>6</sup>

The classic way to change someone's attitude would be to dispute with them, hence, why the two parties would have a reason to dispute. This norm could perhaps still guide our disputing behaviour in cases of intractable disagreement. We are not omniscient; and when it comes to something so unobservable as someone's mental states, frequently there is the epistemic possibility that we could perhaps change these states by disputing with their holder. This is so, even if our disagreement with someone is intractable. Likewise with our motivation to dispute with someone who disagrees in desire with us; even if our disagreement is intractable, the person with conflicting desires to us could still be an obstacle to us getting what we want. Hence, the motivation for changing their attitudes, even in cases where we cannot do this. But as we have seen, disputes in attitude can be pretty much settled without the contested attitudes changing. The settlement concerns what state of affairs shall be realised. (What we are concerned with here is not merely Alan Gibbard's disagreement in plan (2003). Disagreement in plan

---

5 It may not be the case that all disagreement in pro-attitudes takes this shape. Instead of merely wishing to change what the other person will do, a disputant may want to change the other's relevant desires, because they may see them as intrinsically bad to hold. Perhaps John's mother may hold that football is an immoral activity, and so she may see any desire to play it as indicative of a moral failing.

6 This reason would be a *prima facie* one, because of potential factors which could override it.

would stretch much further than the kinds of cases we are talking about. At the moment we are only extending our account to those cases where people have incompatible desires; we are concerned not with ones where the agents' favoured states of affairs are not incompatible. An instance of the latter could be where one person desires candy, and another wants to not have candy. Both of these desires could be jointly-satisfiable )

So it is clear to see how some disagreements in attitude can ground certain disputes: When two people have conflicting desires they want jointly incompatible states of affairs to come about. This type of disagreement in attitude would be most appropriate for grounding disputes over which state of affairs should be brought about. If the disputants argue in this manner when they do not have relevant incompatible desires but think they do, then the dispute is ungrounded; the dispute cannot decide which of two incompatible states of affairs—each favoured by a particular disputant—should be brought about when the disputants do not actually desire incompatible states of affairs.

## 2

## Verbal Disputes

If you have spoken to many evangelical Christians, then you may have heard one say something along the lines of “Christianity is not a religion, it’s a faith” or “I don’t follow any religion, I follow Jesus Christ!” They might even object to someone who refers to Christianity as an Abrahamic religion on the grounds that it is not a religion. If a dispute arose from this misuse of the word ‘religion’ then it would be pointless and superficial. Imagine that someone utters a sentence such as “Christianity is a religion” and an evangelical Christian replies with the sentence “Christianity is not a religion” Our evangelical Christian would clearly have a different idea of what counts as a religion from the person who (correctly) refers to Christianity as one. I suspect that such a dispute would be merely verbal, so that there would not be anything of substance (aside from perhaps the correct use of language) to it. This would explain why the dispute seems so superficial; at a first glance, it looks like the two disputants could perhaps agree on matters of fact (not about linguistic use) and still argue over whether Christianity is a religion. At the very least, this dispute would have features common to many verbal disputes. For instance, the dispute would evaporate if the disputants resolved the linguistic issue (the correct usage of the word “religion”).

Verbal disputes, like this one, would be interesting to philosophers for a couple of reasons. First, they could show us something interesting about language. Second, if we manage to formulate some good criteria for two parties

verbally disputing, then we could perhaps use it to show whether certain disputes in philosophy are verbal. It is taken that verbal disputes are not worth pursuing seriously; the best course of action in a verbal dispute would be to eliminate the terminological confusion. If the dispute is merely verbal then settling the terminology could also eliminate the dispute, or if it is partly verbal then this could show where the real disagreement between the two parties lies. The important thing is that the merely terminological aspect of a dispute should be eliminated—if of course the disputants do not intend to be disputing about language. This implies that if certain philosophical disputes are merely verbal, they should be abandoned. One proponent of this strategy is Eli Hirsch, who argues that there is nothing substantively at stake in questions of the ontology of physical objects “beyond the correct use of language” (2005). Other such disputes in metaphysics which have been taken at some point to be merely verbal include the dispute over whether free will is compatible with determinism and the dispute over whether material objects have temporal parts. In this chapter I shall review some of the main theories of verbal disputes, and discuss some of the features such a theory should have.

## A characterisation of verbal disputes

In the last chapter we distinguished disputes from disagreement. Two people would be having a dispute when they are engaged in some argumentative activity. So that one party would assert a sentence “S” and another would assert “Not-S” in reply, or at least they would reply with something they believe incompatible with “S”. On the other hand, disagreement would be a relation between their mental states, which partly consists<sup>7</sup> in them being somehow incompatible. For

---

<sup>7</sup> I say “partly consists” because, as we saw in the previous chapter, mere incompatibility of mental states may not be sufficient for disagreement. Although I didn't endorse a particular analysis of disagreement, I did argue that disapproval of someone holding an incompatible attitude from oneself was characteristic of disagreement. As I argued, the notion of disagreement also stands in a justificatory relation to disputing. These features of disagreement may arise from incompatible attitudes (in conjunction with our human natures), but they would not be identical to them.

example, two parties' beliefs would be incompatible if one party's were false while the other's were true. Also, two parties' desires would be incompatible when they could not both be satisfied. Although in the previous chapter I advocated the view that we can disagree in our desires, in this chapter I will mainly be concerned with disagreement in beliefs. Sometimes we use the word 'disagreement' to describe the argumentative activity, in fact a few writers refer to this as a species of disagreement (Cappelen & Hawthorne, 2009: 60-1). These disagreements in activity seem to be exactly what I call 'disputes', so I will treat the cases Cappelen and Hawthorne have in mind as such.

For our example of a verbal dispute, let's say that Jenny asserts the sentence "Gordon is wearing pants", and Wolfgang denies it. Despite the linguistic behaviour of Jenny and Wolfgang, they would both agree that Gordon is wearing underwear with no trousers, and they would both be correct in this. They would each agree on what the clothing Gordon is wearing looks like, and furthermore neither would retract their view from further inspection of Gordon and his attire. They are clearly having a dispute; they are both exhibiting argumentative behaviour, etc. But it seems superficial, as if there is nothing grounding the dispute. It should be uncontroversial that this dispute is verbal, what we need is an account on which this dispute comes out as verbal—along with other disputes which are pointless, nonsubstantive, and shallow in the same way.

### *Are verbal disputes about language?*

One idea is that verbal disputes are *about language* as opposed to the apparent first order domain which they seem to be about. At a first glance, an account which relied on this idea would not isolate the class of disputes we want it to. It would rule out verbal disputes where the disputants keep their arguments in first order terms. It would also include disputes which are about language in a way which

would not be especially pointless. Disputes in linguistics and philosophy of language may be about language, but they may not be pointless like the dispute between Jenny and Wolfgang would be. For instance, two people could argue over the meaning of 'pants', because they are trying to build a decent dictionary entry on the word. The disputants in this case would be self-conscious about the verbalness of their dispute, whereas Jenny and Wolfgang would not. The lexicographers would not be too worried by the charge that their dispute is terminological. Furthermore, arguments between lexicographers may be grounded in real disagreement, for example over empirical issues regarding word usage, whereas Wolfgang and Jenny's dispute appears to be first order and so they most likely would not resort to empirical arguments about how we use the word 'pants'.

The disputes we are concerned with in this essay are ones where the verbalness is a defect; in these cases the disputants would not intend to merely argue over words. A verbal dispute may take place completely, in Manley's words, "at the object level" (2009: 12); for instance, Jenny may not believe Wolfgang is mistaken about the meaning of 'pants,' but rather that he was looking at the wrong person, or his vision was failing him, or the sun caught his eyes, and so on. In this case we can say the disputants are each concerned with what Gordon is wearing, not merely with what words they use to describe Gordon's attire. (There are more sophisticated ways of cashing out the idea that certain disputes are about language, which we will consider in a later section.)

So, simply saying that verbal disputes are about language not only includes many disputes which may well be worth pursuing, but also excludes the disputes we are concerned with; namely the ones where the disputants are unaware that their dispute boils down to the use of words. The cases we are concerned with are the ones where the disputants do not know the full extent of their dispute's verbalness. (These may even encompass disputes about language, for instance, David Chalmers suggests disputes over the semantic-pragmatic distinction may be verbal (2011: 532).) We need an account to cover cases which are not about language, or at least are more verbal than they seem. Especially so when we factor in the second reason for studying verbal disputes; that certain philosophical disputes may be verbal without us knowing.

### *Do the parties to a verbal dispute disagree?*

Another putative explanation of why disputes like the one between Jenny and Wolfgang are verbal is that the disputants do not disagree; they just talk past one another. By this I mean that they do not have conflicting non metalinguistic beliefs<sup>8</sup>. Sure, they may disagree on matters of language, for instance, on what word to use to denote such-and-such item of clothing; but they would not disagree on nonlinguistic matters. Now, this explanation is too simple. For a start, it is clear that two parties can be involved in a verbal dispute, and yet still disagree on certain matters. The verbalness of Jenny and Wolfgang's dispute would not be affected by them disagreeing on what Gordon's birthday is.

The definition may also exclude a great number of disputes in the way it explains the disputants' lack of disagreement. One may say, for example, that the sentence "Gordon is wearing pants" expresses different propositions in the respective mouths of Jenny and Wolfgang; Jenny would express that Gordon is wearing underwear, but Wolfgang would express that Gordon is wearing trousers. In this way, Jenny's utterance of the sentence would be compatible with Wolfgang's utterance of its negation; surely Jenny saying Gordon is wearing underwear is compatible with Wolfgang saying that he is not wearing trousers.

It is not clear that the sentence "Gordon is wearing pants" *wouldn't* mean different things in each others respective mouths. Hilary Putnam argued that the meanings of natural-kind terms "just ain't in the head," they also depend on the speaker's environment. Others, like Tyler Burge, have since extended this thought to other kinds of words and to the content of psychological states. Burge (1979) argues that someone's linguistic environment—their linguistic community—

---

<sup>8</sup> I take it that disagreement is best cashed out in terms of two parties belief content. It neither depends on what they say conflicting, or even on the two parties having a dispute. One may be dishonest and actually agree with the other party, and I may disagree with someone who I have never met.

determines what their words express. If Jenny and Wolfgang are both part of a linguistic community where 'pants' denotes trousers, then one of them (Wolfgang) just says something false. Not only that, but Burge would say that the *belief* Wolfgang expresses when he says "Gordon is not wearing pants" is false. In this way Jenny and Wolfgang *would* perhaps disagree in some sense, as they would believe incompatible propositions. It is in this way that a definition of verbal disputes could not simply say that the disputants do not disagree over nonlinguistic matters and then leave it at that. Accounting for externalism—especially the social externalism of Burge—is necessary for a theory of verbal disputes.

A verbal dispute is then not characterised by having no disagreement present at all between the disputants, or by having no disagreement over the subject matter of the dispute. But this is not to say that our notion of disagreement is unimportant for the investigation into verbal disputes. The best strategy, I take it, is something rather like what Chalmers presents in *Verbal Disputes*, where disagreement is allowed, providing the dispute (and presumably, any disagreement over the subject matter of the dispute) arises from a certain source. I will detail Chalmers's account later on in the chapter, but for now let us note that his placing importance of where a dispute arises from is something echoed by the previous chapter. In fact, a (very vague) way of stating Chalmers's thesis is that a dispute is verbal when it is not grounded by the right kind of disagreement.

### *Verbal dispute internalism*

A feature of verbal disputes relating to the issue of semantic externalism is that the verbalness of a dispute depends more on factors intrinsic to the disputants than to their environment. For almost any verbal dispute we can imagine a counterfactual scenario where a social externalist would say that the disputants

(or intrinsic duplicates of them) disagree (Chalmers, 2011: 520). For example, above we considered the case where, rather than coming from separate linguistic communities which use 'pants' differently, Jenny and Wolfgang could come from the same community with a single fixed meaning for the word. This counterfactual dispute would still be verbal; it would suffer from the same defects as other verbal disputes.

We can perhaps extend this thought further, so that when two parties A and B are having a verbal dispute, then if intrinsic duplicates of them, A1 and B1, had the same dispute—argued with the same words, the dispute arose in virtue of the same divergence in linguistic usage, etc.—then A1 and B1 would be in a verbal dispute whatever environment they were in. We can call this verbal dispute internalism (VDI). This is unlike the case of mental content where the externalist would say that a person's belief could have different content depending on their environment. The interesting thing to note here is that we can accept the externalist picture of mental states and VDI.

A few provisos here. First, the disputants' doppelgängers would be intrinsic duplicates with the same physical constitution and the same (or at least very similar) brain activity at any given moment. As a result of this, they would be disposed to behave identically, in that they would make the same motions, say the same sentences, etc. So there is a natural restriction on the environment here, i.e. the counterfactual environment must be such that the disputants couldn't tell the difference between the two environments, the disputants wouldn't behave differently, etc. Second, this thesis is *not* that the state of being in a verbal dispute is not relational. In fact I would say it is essentially relational: Agreement and disagreement is the relation between multiple people's mental states, and disputing is a relation between two people where they attempt to resolve disagreement. However, it does not matter too much which environment the disputants are related to each other in, for the dispute to be merely verbal.

This has something in common with the view of mental content Jackson and Braddon-Mitchell present (2007). They say of content that “We understand it, at least in part, in terms of interactions with things around us—inputs and outputs described in distal terms” but they add that “it does *not* follow that

content is not narrow in an important sense” (215). They state “The narrowness of narrow content is constituted not by its being independent of interactions with, and relations to, the environment, but rather by the fact that which interactions are actual and which possible does not matter” (216). They see narrow content as reflecting the way a subject would behave, etc. in certain situations; but the way a subject would interact with an environment is determined by their internal nature, and their internal nature may cause them to act the same way in perceptually indistinguishable, but nevertheless different, situations.

The nature of verbal disputes is analogous to Jackson and Braddon-Mitchell's view of content in this way: A verbal dispute is verbal in virtue of the subjects' mental states and their relation to one another. They are also relational in virtue of mental content (which most would agree is relational) playing a key role in the explanation of verbal disputes. However, although verbal disputes involve essential reference to the subjects' interactions with one another and their environment, what the actual environment is does not matter. Jenny and Wolfgang's internal states and their argumentative behaviour together entail that in almost any environment their dispute will be verbal. I say 'almost any environment' because there may be possible situations where the natural laws are different from the Jenny and Wolfgang's world which may mean that they are not involved in a verbal dispute.

Against this position, one could perhaps formulate an externalist account of verbal disputes. This account would say that the environment *does* make a difference to whether a dispute is verbal or not. So that Jenny and Wolfgang would be in a verbal dispute when they are from separate linguistic communities. But in the case of them belonging to the same community, they would not be in a verbal dispute because of their shared language. As we have seen, this would perhaps not be a good move. Chalmers points out that saying a dispute like this is not verbal because the disputants do not disagree—do not express different propositions by uttering the disputed sentence—misses the point. The two cases feature the same kind of triviality and so our explanation of verbal disputes should encompass both. If we find out that the latter case features disagreement then the correct response would be to say that non-disagreement is not a

necessary condition for being in a verbal dispute, rather than proclaiming that the dispute cannot be verbal. Thus, we should be internalists about verbal disputes.

## Hirsch on verbal disputes

Eli Hirsch has been one of the more prolific writers when it comes to verbal disputes. His view of what a verbal dispute is evolves throughout his work, however a theme running through all of it is the idea that certain disputes in metaphysics are merely verbal. I shall only cover two different definitions; one from *Dividing Reality* and another from his later work (2005, 2009) in terms of imagined linguistic communities.

In *Dividing Reality* Hirsch states that a dispute is verbal if it meets the following conditions:

*The Equivalence Condition:* For any controversial sentence S within the dispute, there are two sentences Sa and Sb (the “reconciling sentences”) such that (i) neither Sa or Sb is controversial, and (ii) one disputant believes that S is equivalent to Sa and the other believes that S is equivalent to Sb (1993: 181).

*The Consistency condition:* Each disputant’s position with respect to the controversial sentences is consistent with what the disputant would say after further observation or argument, discounting empirical arguments about the conventions of language (1993: 182).

(As when talking about other verbal disputes, I assume that the two disputants each are asserting seemingly contradictory sentences, e.g. “S” and “Not-S”.) The consistency condition is there to rule out cases of inadequate reflection. For example, someone may take the word 'vehicle' to only apply to land vehicles like cars and lorries, because they don't associate the term with air-vehicles (e.g. planes) or sea-vehicles (e.g. boats). But if they were to consider and reflect on these cases then they would come to realise that planes and boats are vehicles too.

Note that because it discounts “empirical arguments about the conventions of language”, the definition would still include cases where a semantic externalist would say the disputants do not disagree. In the case where they both speak British-English, Wolfgang may well accept that Gordon is wearing pants if Jenny says “In Great Britain we use the word 'pants' to refer to trousers” and points out that he is diverging from established usage. Jenny and Wolfgang's dispute seems to meet this condition. If all of Jenny's arguments are for a position consistent with holding that Gordon is wearing trousers, and Wolfgang really does take 'pants' to refer to trousers, then any argument (except ones concerned with the use of language) for Jenny's position will be consistent with Wolfgang's.

The pants case would also meet the equivalence condition. Let's call the sentence “Gordon is wearing pants” S. Jenny would believe it equivalent to “Gordon is wearing underwear” and Wolfgang would believe it equivalent to “Gordon is wearing trousers.” Neither of the latter two sentences would be controversial for either disputant. So at a first glance seems that Hirsch's (early) definition covers this case.

However, as Chalmers points out, if a dispute met Hirsch's conditions then the dispute would indeed be verbal, but the conditions wouldn't be necessary. For a start, the disputants may not take each respective reconciling sentence to be equivalent, in a strict sense, with the sentence S; strict equivalences seem hard to come by in real life (2011: 519). Another problem for this characterisation is that there may be no reconciling sentences. Perhaps the term 'underwear' is not in Jenny's vocabulary; she may not believe the word 'pants' to be a synonym for any other word or complex expression. She may have learnt the word by ostension, so her only criterion for something being a pair of pants is that it be 'something like *that*' (when gesturing at a pair of pants). In this way we could not formulate reconciling sentences for her and Wolfgang's dispute.

These worries may be part of the reason why Hirsch characterises verbal disputes differently in later work. He says “The general characterisation of a verbal dispute is one in which the controversial sentences are most plausibly interpreted as having different truth conditions, so that each position turns out to be correct in its associated language” (2005: 72). So if the disputants were just

speaking in different languages—or rather different idiolects—they wouldn't contradict each other even though they appear to. This may appear to directly contradict social externalism, but Hirsch's account is made more nuanced by his appeal to imagined linguistic communities<sup>9</sup>. David Chalmers gives us an accurate summary of Hirsch's second characterisation:

Two parties A and B are having a verbal dispute iff, were A and B to inhabit an A-community and a B-community (respectively) in which everyone exhibits the same sort of linguistic behaviour that A and B actually exhibit (respectively), then on the correct view of linguistic interpretation, A and B would agree that both speak the truth in their own language. (From Chalmers, 2011: 521, paraphrased from Hirsch, 2009: 238–40.)

According to Hirsch, the correct view of linguistic interpretation is the assignment of truth conditions based on the principle of charity, that is, we try and interpret the parties' respective linguistic behaviour as being maximally truthful and rational (2005: 71). This initially does not seem to help much, as the correct view of linguistic interpretation may have to take into account the deferential dispositions of the parties. If Wolfgang is a British-English speaker then he will likely defer to established usage in Britain, so according to social-externalism, he will be most plausibly interpreted as saying something false. However, things seem different when we imagine two linguistic communities whose respective usage roughly corresponds to the usage of Jenny and Wolfgang; so we would have one community which would use 'pants' to refer to underwear and another which would use the same word to refer to trousers. We could call the former's language J-English and the latter's W-English, so that the sentences Jenny asserts about Gordon would be true in J-English and Wolfgang's in W-English. If Jenny inhabited the J-English community and Wolfgang inhabited the W-English community then they would not disagree over the subject matter of the dispute; they should both agree that they would speak the truth in their respective languages.

This characterisation can account for cases where one disputant does not

---

<sup>9</sup> Of course, in the case of Jenny and Wolfgang actually belonging to different linguistic communities we do not need to imagine them, as they already exist in the form of British and American English.

believe that S is equivalent to another non controversial sentence. Hirsch states that the same sentence S would plausibly be interpreted as having different truth conditions in different languages. This would be in the same way that phonetically and typographically identical expressions in different languages may have different meanings, for instance, the words 'sensitive' and 'bar' have different meanings, respectively, in Italian and English<sup>10</sup>. The sentence S having different truth conditions in J-English and W-English would be entirely compatible with the disputants not believing S to be equivalent to any other sentence not involving the word 'pants'.

A potential problem with Hirsch's later characterisation is in the way it is supposed to be compatible with social-externalism. We can see how Hirsch takes it to be so from the quote below:

If Burge's view applies to this example, it implies that A is not speaking (or thinking in) A-English (or in a corresponding private idiolect); rather A has the mistaken thoughts and beliefs that are expressed by her assertions in plain English. I'm not entering into that question. *What is important for my purposes is that the sentences asserted by A are true in A-English, so that the only real question is whether A-English is plain English. This is why the dispute with A is merely verbal.* In effect I am redefining “A's idiolect” to mean the (imagined) public language associated with A's position. This redefined sense of “A's idiolect” captures the relevant sense in which “A is right in A's idiolect (and we are right in ours)” (2005: 70 (my emphasis)).

(Here A is someone who believes that glasses count as cups. A-English identical to plain English except for the word “cup” also applying to glasses. So the A-community would be a group of people who spoke A-English.)

From the emphasised part of the quote we can see that Hirsch interprets the question over whether the glass is a cup (in the context of his example), as a question *about language*; presumably it is something like 'Does the word 'cup' also denote glasses?' and the obvious answer—because Hirsch says that the disputants both claim to be talking plain English—is whatever is correct in plain English (or rather, the language the disputants speak). This is why he appeals to separate

---

<sup>10</sup> Thank you to Ambra D'Antone for the example.

linguistic communities; if the question is about what language plain-English is then the question could be formulated as “Is plain English A-English (where glasses count as cups) or B-English (where glasses do not count as cups)?”

But if we leave it at that, the account may face the difficulties raised for accounts which describe verbal disputes as being about language (Manley, 2009: 12). So, perhaps by being 'about language' Hirsch means that the disputes are *covertly* about language. But what would this mean? If the dispute does not appear to be about language—e.g. if it was formally identical to a nonlinguistically substantive dispute—then what would this aboutness consist in?

Perhaps we could say that a dispute is covertly about language just in case it could be dissolved by reflection on language; if we solved the linguistic issue then the whole dispute would (hopefully) disappear. This would be consistent with Hirsch's statement that the only “real question at issue is which language is plain English” (2005: 70). Once we see which language is plain English and adjust our linguistic behaviour accordingly, then we have resolved the dispute. Also see how Hirsch's use of linguistic communities fits into this. If the only relevant question is which way of speaking is the correct way, then the question can be rephrased as asking which linguistic community most correlates to ours; should we speak like the A-community or the B-community?

This strategy may well work. But notice that it would very much weaken the claim that a dispute is about language. The way it is about language corresponds more to the way in which we could resolve the dispute or what actual issue is in the vicinity, rather than the natural way which would be the disputants consciously trying to address a metalinguistic issue. This does not matter very much at all though; the important thing to notice is that it shows how the only real question in a verbal dispute is what language is the correct one to use.

Now a problem that Chalmers raises for Hirsch's (later) definition is that it cannot account for certain other cases of semantic externalism; ones where the deference to the linguistic community is not at issue in the dispute. Chalmers mentions disputes where inadequate reflection is the source of the dispute such as this:

Suppose that A and B agree that Sue made a false statement that she did not believe to be false, and also agree on the moral status of Sue's assertion and other relevant properties. A says 'Sue did not lie'. B initially says 'Sue lied', believing falsely that 'lie' refers to any false statement, but on reflection comes to accept 'Sue did not lie', through reflection on the concept of lying . . . But, they are having a broadly verbal dispute all the same: intuitively, they agree on the important facts of the case and are merely disagreeing on whether the word 'lie' should be used to describe it (2011: 520-1).

Why could Hirsch's characterisation of a verbal dispute not accommodate this case? The reason is that it would be strange for there to be a B-community if B does not even have a proper handle on her concepts. We can see this with the example of a pre-Gettier philosopher who argues that knowledge is justified true belief. This philosopher may argue that the concept they possess is JTB, they could even describe Gettier scenarios as cases of knowledge; but they may realise that they have been applying the term incorrectly, perhaps because they realise knowledge is, for example, reliable by definition and JTB doesn't give us this result.<sup>11</sup> In the same way, B may realise that applying the term 'lie' in the way that they do, is irrational; for instance, lying may have some kind of moral implications, which B finds hard to balance with her belief that Sue did not do anything morally wrong. This is not to say that there could not be a B-community where everyone takes "lie" to denote any false statement. This community would just not represent B and the meaning she attaches to 'lie'. Furthermore, because B is irrational in their employment of the disputed term, we may not—or rather we should not—interpret their statements charitably. Thus, Hirsch's later definition would have trouble accounting for this case. Even Hirsch's early definition would face a similar problem because the consistency condition entails that the disputants' positions are consistent with what they would say after further reflection. B's utterance "Sue lied" could be dissolved by this reflection, so they would not meet the consistency condition. But this case would still clearly be a verbal dispute.

---

<sup>11</sup> Thank you to Jessica Brown for the example.

## Chalmers on verbal disputes

What we learnt in the last section is that we need a theory of verbal disputes to cover cases where each disputant may not believe that different sentences are equivalent to the disputed sentence. We also need it to encompass disputes which are not obviously *about* language, and ones where the disputants may have contradictory beliefs as a result of one disputant's semantic deviance. Chalmers recognises this and so his definition gets over the worries we had about Hirsch's definitions. Chalmers presents what he calls the narrow conception of verbal disputes:

A dispute over *S* is narrowly verbal iff *S* expresses distinct propositions *p* and *q* for the two parties, so that one party asserts *p* and the other denies  $q\sim$ , and the parties agree on the truth of *p* and  $\sim q$  (2011: 519).

Now the narrow conception would encompass disputes between different linguistic communities, for example, it would cover the dispute between Jenny and Wolfgang if the former was from Britain and the latter was from America. However, it would not cover the dispute if both disputants were part of the same linguistic community as then they would both defer to, say British-English usage. In the latter case an externalist like Burge would say that the sentence “Gordon is wearing pants” would express *the same* proposition for each disputant; namely, whatever proposition the usage of the sentence for the relevant linguistic community expresses. The correct reaction to the latter case would not be to say that such a dispute is not verbal, but rather to say that it is a particular type of verbal dispute, namely what Chalmers calls a *broadly* verbal dispute:

A dispute over *S* is broadly verbal when, for some expression *T* in *S*, the parties disagree about the meaning of *T*, and the dispute over *S* arises wholly in virtue of this disagreement regarding *T* (2011: 522).

The broad characterisation—which also encompasses narrowly verbal disputes—would include those disputes where the disputants may not express different

propositions with a sentence S, but which have the same faults as the disputes which do. Jenny and Wolfgang clearly disagree over the meaning of 'pants,' and this disagreement plays a key explanatory role when we account for this dispute; the dispute would not arise at all if it were not for Jenny and Wolfgang's divergent uses of language. This key explanatory role is what Chalmers identifies with the 'in-virtue-of' relation between the metalinguistic disagreement and the dispute (2011: 523). It is the broad definition I shall be concerned with when discussing Chalmers in the following sections unless otherwise specified.

Chalmers's broad definition of verbal disputes clearly has advantages over both of Hirsch's. For a start, unlike Hirsch's first definition, it would not require both disputants to believe in an equivalence between the disputed sentence and two different respective sentences. Jenny could have no other word for underwear other than pants, but still disagree with Wolfgang over the meaning of 'pants', perhaps in the way she would apply the word across possible cases. Also, Chalmers can account for disputes which are not about language (in the strong sense). Metalinguistic disagreement can explain a dispute without the dispute being about language. A dispute can take place entirely at the object level whilst the only real disagreement is metalinguistic. For instance, the dispute over whether Gordon is wearing pants could be first order—with the disputants perhaps criticising each others' vision—whilst the metalinguistic disagreement gives rise to the dispute. However, using metalinguistic disagreement to define verbal disputes may cause some problems for Chalmers, as I shall explain below.

In the next few sections I shall look at where some potential opposition to Chalmers could come from. I aim to defend Chalmers's account (or one very much like it) from these objections. I will suggest some positive changes, in line with what other writers like Herman Cappelen and C. S. I. Jenkins say, but these will not affect what I think is the key insight of Chalmers's account, which is that verbal disputes are those which arise in a certain way. This thesis has certain similarities to what I present in the first chapter; there I argue that we accept norms entailing that a dispute should be grounded by the right kind of disagreement.

### *Wholly, Partly, and Merely Verbal Disputes*

Chalmers importantly talks about *wholly* verbal disputes as opposed to *partly* verbal disputes. A wholly verbal dispute would fit his definition of a verbal dispute in that the dispute arises wholly in virtue of the meaning of some expression. On the other hand a partly verbal dispute is one where the “apparent first order disagreement arises partly in virtue of the metalinguistic disagreement and partly in virtue of a substantive non-metalinguistic disagreement” (2011: 525-6).

The subject we wish to engage with is closer to the wholly rather than the partly verbal dispute. The reason for this is that, generally, it would be a far more damaging for a dispute if it was wholly instead of partly verbal. Philosophers who take metaphysical disputes to be verbal usually take it to be a problem with the subject matter, so that there really would be nothing more to the dispute than choosing a particular language. If a metaphysical dispute is partly verbal, then this means that there could be still something substantive at issue in the dispute, only perhaps not what the disputants originally thought was at issue. In this way, a dispute being partly verbal would not warrant deflationism about the subject matter of the dispute. This is not to say that showing a dispute to be partly verbal is a useless result. Chalmers for example, states “the diagnosis of verbal disputes has the potential to serve as a sort of universal acid in philosophical discussion, either dissolving disagreements or boiling them down to the fundamental disagreements on which they turn” (2011: 517). In this way, if a dispute is wholly verbal it will be dissolved. If it is partly verbal, it will be boiled down to the fundamental disagreement.

There is a further distinction to be made here, one which we have already alluded to. It is the one between *wholly* and *merely* verbal disputes. Although Chalmers mentions merely—as opposed to wholly—verbal disputes, he acknowledges Jenkins in a footnote about it, and Jenkins has a later paper about the distinction (Jenkins: 2014).

In informal terms, the difference between a merely and a wholly verbal dispute is that the former is one where nothing non-linguistic is at issue but where it appears as if there is. A wholly verbal dispute would be one where the issue is verbal but the speakers do not necessarily intend that there be anything non-linguistic at issue. An example of a wholly verbal dispute could be a situation where the participants are arguing over the meaning of a particular word. There would not clearly be anything non-linguistically substantive at issue in this dispute; this would be fine with the participants as they know they are arguing about language. In this way merely verbal disputes would be a subset of wholly verbal disputes; the set of wholly verbal disputes in which there is an appearance of non-linguistic disagreement in the surface dispute. We already touched upon this distinction in the section on Hirsch and noted that accommodating *merely* verbal disputes, the ones which are not about language, was a potential problem for him. Jenkins's definition of a merely verbal dispute is as follows.

MVD: Parties A and B are having a merely verbal dispute iff they are engaged in a sincere *prima facie* dispute D, but do not disagree over the subject matter(s) of D, and merely present the appearance of doing so owing to their divergent uses of some relevant portion of language (2014: 21).

Jenkins intends for merely verbal disputes to be wholly verbal in Chalmers sense, in that in the disputes are not partly verbal and partly non-verbal. But she also wants to make the further claim that in merely verbal disputes there is the *appearance* of the dispute being about something non-linguistically substantive. She attempts to capture those verbal disputes which seem inappropriate to carry on with once it has been pointed out there is nothing substantive at issue (2014: 11). This is an important point; when a philosophical dispute is dismissed as verbal, it is usually taken as a given that the disputants do not know that the dispute is verbal. Often disputes in metaphysics are the target of the verbal dispute strategy. These disputes are typically taken (by metaphysicians) to be fundamental and substantive; so the charge that these are merely verbal disputes would be unwelcome to a thoroughgoing realist about the subject matter. Disputes such as these do not *prima facie* appear to be about language (to the disputants at any rate) and so the charge that they are verbal is far more damaging to them than a dispute over, say, the correct pronunciation of

'Worcestershire'. The charge that a dispute is partly verbal does not have the same force. The implication is that if a dispute is partly verbal there may be something substantive at issue, and that we just need to rephrase the dispute so as to avoid speaking past one another. I will not go into detail about Jenkins's theory in general, the moral of this section is that an account of verbal disputes should be able to intelligibly isolate the disputes which are *merely* verbal.

Can Chalmers's account do this? I believe it can simply with an extra condition stating that merely verbal disputes are the ones which meet his definition of a 'broadly verbal dispute' and where "usage of the key term is not itself a central object of concern" (2011: 525).

### *Metalinguistic disagreement*

One objection Chalmers anticipates is that metalinguistic disagreement may not play any key explanatory role in verbal disputes at all; either because users of a word do not need beliefs about its meaning, or because the metalinguistic disagreement may not give rise to the apparent first-order disagreement (it may be the other way around). Let's use our comfortable and familiar example to illustrate this objection, in particular the former version. Jenny and Wolfgang may not need any beliefs about the word 'pants' in order to be having a verbal dispute over whether Gordon is wearing them. As Herman Cappelen points out, a speaker may use an expression E, and have beliefs about E, without having any beliefs about the meaning of E. Cappelen states that he doesn't think "there's an interesting subset of beliefs of the form 'Es are G' that are beliefs about the meaning of 'E' and a theory of verbal disputes shouldn't require that we be able to make that distinction" (2013: 44). Perhaps Jenny could believe that pants are an item of clothing which typical people wear under trousers and superheroes wear over trousers, without believing anything of *the word* 'pants.'

I will keep out of the debate concerning whether there is such a subset of

beliefs, but if there isn't then it would perhaps be fatal to Chalmers's account as it stands. Inspired by Jenkin's account, Chalmers could perhaps refine his account so that he talks about the disputants' divergent uses of language instead of metalinguistic disagreement. Then his broad account would look something like this:

A dispute over S is broadly verbal when, for some expression T in S, the parties use (or are disposed to use) T in different ways, and the dispute over S arises wholly in virtue of this divergence in the usage of (or the disposition to use) T.

We could also stipulate that, not only the dispute, but also any disagreement over the subject matter of the dispute would arise from the divergent uses of language in a verbal dispute. So that in the case where Jenny and Wolfgang inhabit the same linguistic community, their disagreement over whether Gordon is wearing pants, arises wholly from their dispositions to use language, than from their dispositions to behave otherwise.

This would be to block cases where the dispute arises in virtue of the divergent uses of language, but where, independently, there is relevant disagreement. Examples of this could be seen in disputes between experts and novices. For instance a white-belt in Judo could say, while watching a competition “Wow what a stunning O goshi” and their sensei may reply “No that was a stunning *uki* goshi.” The dispute clearly arises because the novice used language differently from the expert—if the white-belt had used the correct term then the dispute wouldn't have occurred—but there would be disagreement between the two, which would not arise in virtue of their divergent uses of language. The two throws look very similar, in fact so similar that it would probably take a relatively experienced judoka to tell the difference between them—the main difference being that in performing an O goshi you sit lower and use a pulling motion, as opposed to an uki goshi where you use a twisting motion. The novice judoka may know that in an O goshi you pull your opponent over your hip (it is often the first throw you learn) but may not actually know that there is a similar throw called 'Uki goshi' where it doesn't. As a result of their untrained eyes, the novice could perhaps think that the victor of the Judo match did pull their opponent over their hip. In this case the disagreement between the novice

and the sensei could be cashed out in terms of whether they think the competitor in question threw their opponent over their hip or not.

### *The in-virtue-of relation*

A very important aspect of Chalmers's broadly verbal disputes is the *in-virtue-of* relation; a dispute over S would be broadly verbal if it arose in virtue of metalinguistic disagreement over the meaning of some expression in S. Now, Chalmers does give us a hint at what he means by 'in-virtue-of.' He states:

Here, as elsewhere in philosophy, one may have to take something in this vicinity as primitive, but one can at least offer some clarification. I think this should be understood as an explanatory “in virtue of”: the idea is that the metalinguistic disagreement explains the apparent first-order disagreement. And the relevant sort of explanation should require something stronger than an arbitrary causal or evidential explanation: it would not suffice, for example, for a metalinguistic disagreement to cause personal enmity that then causes first-order disagreement (2011: 523).

This sounds slightly mysterious, so it is not clear whether or not any criticism (apart from mystery) will stick. As is clear from the quote, a counterexample where Jenny and Wolfgang's metalinguistic disagreement over 'pants' leads to them being at loggerheads, which itself causes their dispute over who's turn it is to take out the dustbin, would not be on the right track at all. Nevertheless, I will present a potential flaw with the definition.

One type of case which this definition may have trouble with where the dispute arises wholly in virtue of metalinguistic disagreement over the disputed sentence S, and where there is still relevant, non-linguistic disagreement. Take for example a dispute between Jeff, who asserts the sentence “there is beer in the fridge” and Sheila, who denies it. Let's say that the dispute arises because Jeff takes cheap lager to be beer and Sheila does not; Sheila is the more discerning

drinker of the two. They both sit in Jeff's living room where they discuss different types of alcohol, and it comes out that Jeff only buys cheap lager. He then later on says that he could get a beer from his fridge for them both. Sheila, however, corrects him by saying that "there is no beer in your fridge" because she does not take lager to count as beer and she knows that Jeff only buys lager. However, a further detail to this case is that, unbeknownst to Jeff, but known to Sheila (she had a peek whilst nabbing a yoghurt earlier in the evening), there is no lager (or ale) in the fridge (let's say that this fact wasn't in her mind when she corrected Jeff). So the objection would go, even though in this case the dispute arose wholly because of the metalinguistic disagreement, the dispute is not merely verbal because the disputants disagree on a relevant matter, i.e. the contents of the fridge.

Now as shown, some disagreement is permissible in a verbal dispute—in fact Chalmers's definition is designed to accommodate it—but it must be the right kind of disagreement. We would allow irrelevant disagreement, such as Jenny and Wolfgang's relative stances on why a raven is like a writing desk; or disagreement as a result of semantic deviance, like Jenny and Wolfgang's disagreement on whether Gordon is wearing pants, when they are part of the same linguistic community. Now the disagreement in this case is not of either type. Unlike the former, it is relevant to the dispute (it expresses the disagreement between Sheila and Jeff over the sentence "there is no beer in the fridge"), and unlike the latter it is relatively independent from the way the disputants would apply the disputed term across possible cases. Perhaps an account such as Jenkins's, would do better. Recall, she says that verbal disputants (or in her terms, 'prima facie disputants') do not disagree over the subject matter of the dispute. This would seem to exclude the case of Sheila and Jeff, as they clearly disagree over a relevant matter, i.e. whether the fridge is bare or not.

I don't think this objection would work. One of the main ideas in the last chapter was that our practice of disputing suggests strongly that we view disputes not standing in the appropriate relation to disagreement as unjustified; for us, the right kind of disagreement reflects the subject matter of the dispute and grounds it. The reason for this, I argue, is tied to the default purpose that disputing serves for us, i.e. it is supposed to resolve disagreement. I am not saying here that

disputing intrinsically has this purpose, nor am I saying that it can be used for no other purpose. What I am saying is that our practice of disputing seems to suggest that we accept that disputing is 'for' resolving disagreement.

Take this example: one person may dispute with another with the intention to annoy them. The purpose that the first person has in mind does not require that their disagreement be resolved; in fact it does not even require that the disputants disagree at all. However, disputing can be appropriate for bringing about this intended result because of the parties' conventional understanding that the primary purpose of disputing is to resolve disagreement. If one person sincerely tries to resolve disagreement in favour of their view, the implication is that they take their beliefs to be true and their opponent's to be false. But as we saw, we view false beliefs as not merely false but also *incorrect*, and so apt for being changed. In this way, someone may be annoyed at being told they have false beliefs; they are essentially being told their belief system is defective, for believing false things is a flaw (of course we all have some false beliefs, but nobody's perfect). Of course, this may only work on a certain type of person, e.g. one who places a lot of value on having true beliefs. But we can at least see *why* they would place such value on this, and why they would be offended at being told their beliefs do not meet this standard ("Who are you to tell me *I* am wrong?").

My account in the last chapter is relevant to a defence of Chalmers' definition of verbal disputes because a consequence of it is that not all disagreement will ground a dispute in the right way. This is because not just any old dispute can resolve any old disagreement. For example, the subject matter of the disagreement needs to be relevant to the dispute in order for the dispute to be substantive; disagreement over who to vote for at the next election would be unlikely to be solved by arguing over how to cook brisket for. This point is pretty much in accordance with Jenkins's account, but I would go further and say that a dispute should also arise in virtue of the right kind of disagreement. It is hard to imagine cases where two people disagree over some proposition *P*, argue over its truth, but that their dispute does not arise in virtue of it. We need to posit some mechanism by which the disputants can disagree over the subject matter of the dispute, without their disagreement coming into play. This verges on the absurd.

It is almost incomprehensible that two people could argue over the truth of “P” whilst disagreeing over the truth of P, without letting those beliefs affect the dispute. Likewise, it is not at all clear whether disputes like the one between Sheila and Jeff are really coherent. However, let's allow for a moment that the case of Sheila and Jeff does make sense. I argue that even if it does, it would still not be a substantive dispute; that is, not until Sheila and Jeff's first order disagreement kicks in, but then it would either be a different dispute, or would not be wholly explained by their linguistic divergence from one another.

We can see this from the arguments that they would sincerely employ in support for their positions. Let's say they only employ arguments which arise from their linguistic divergence. Jeff may say something like “There is beer in the fridge, I bought some yesterday” and Sheila may say “But you didn't buy *beer*, as you only buy cheap lager” in reply. In this case, their dispute looks very much like a verbal dispute. They are for all practical purposes speaking past one another, as they both are using different criteria for beer being in the fridge, and the situation which Jeff takes to be the case—cheap lager being in the fridge—would meet Jeff's criteria, but not Sheila's.

However, if they use arguments which spring from their substantive disagreement, then their dispute will fail to operate as a merely verbal dispute. If Sheila instead replies to Jeff with “Well I went to the fridge to nab a yogurt some time ago but I saw it was completely empty” then she will have expressed something incompatible with what Jeff believes, but not as a result of their linguistic divergence.

In this case, the dispute changes from being merely verbal to there being something substantive at issue, namely, whether there is anything at all in the fridge. Of course, there is still the confusion over whether there is 'beer' in the fridge, but there would be some substantive disagreement underlying in the dispute. In this case, their dispute is only partly verbal, rather than merely verbal. We can see in the case of Sheila and Jeff, that they do disagree about relevant matters, but their dispute is in no way grounded by this disagreement; it arises from something other than the actual disagreement. I take it that this case, if the disputants didn't let their actual disagreement come into play, then it would

remain a verbal dispute. If they did let it come into play, then it would cease to be a verbal dispute.

What I have tried to do in this section is to explain why Chalmers is correct in taking it that it is relevant to whether we have a verbal dispute, whether or not it arises from the linguistic differences of the disputants or not. This is opposed to Jenkins, who holds that whether we have a verbal dispute is more to do with what the disputants disagree on, rather than whether or not the dispute arises in virtue of such disagreement.

## Conclusion

I have pointed out some problems with Chalmers's account of verbal disputes, and suggested some amendments. I have also defended what I take to be a key insight of Chalmers's definition: that verbal disputes are those which arise in virtue of linguistic divergence. All things considered, this is the definition of verbal disputes which I favour:

A dispute over S is broadly verbal when, for some expression T in S, the parties use (or are disposed to use) T in different ways, and the dispute over S arises wholly in virtue of this divergence in the usage of (or the disposal to use) T.

This gets over the problem of the disputants not having any particular metalinguistic beliefs, and it is easily targeted at merely verbal disputes by specifying that the disputants intend to be speaking about some first order subject when their dispute does not arise in virtue of any disagreement over this.

## 3

# Conceptual Analysis and Explication

## Introduction

Conceptual analysis is usually taken to be the practice of spelling out the application conditions for our concepts, or in other words: finding out when it is correct to apply a particular concept. It is, as Jackson points out, essential for the project of conceiving of the world in a limited amount of more or less fundamental ingredients. If we can completely describe the Xs in terms of a more fundamental vocabulary then we need not posit a separate ontological category for them; they would be nothing over and above matters described in the more fundamental vocabulary. This seems to give us a framework allows us to solve certain philosophical problems. If we can show that the existence of Xs are entailed by matters described in a certain way then, if we can show that the description is true, we have secured the existence of Xs. Conceptual analysis would be necessary for doing this for Jackson because it is, at its heart, “addressing when and whether a story told in one vocabulary is made true by one told in an allegedly more fundamental vocabulary” (1998: 28). This Chapter addresses Williamson’s inductive argument against conceptual analysis, and what I call the ‘deflationary argument’, which entails that engaging in conceptual analysis is pointless in a similar way to engaging in verbal disputes. I argue that this criticism only works against a particular type of conceptual analysis, and that another type, specifically Carnapian explication, can certainly escape it. In this

way the argument would not be fatal to conceptual analysis in general, only a particular form of it.

With regards to metaphysics, most commentators have focussed on Carnap's overall deflationary project. However, his notion of explication is something that we can take on without committing to Carnap's overall project. This is what I intend to do by showing that Carnapian explication can help us with questions of ontological reduction. I first outline some contemporary views on conceptual analysis, most importantly Frank Jackson's 'Canberra plan' style. I then outline Williamson's argument against conceptual analysis and how it affects the Canberra plan as conceived by Jackson. I will do the same for the deflationary argument against conceptual analysis. I then outline Carnapian explication and how it helps overcome both arguments against conceptual analysis. Finally I address certain responses that could be used against my proposal.

## Conceptual analysis and the Canberra plan

What is conceptual analysis? In order to give an adequate characterisation it will be useful to give some historical background. It is taken by most philosophers that there exist *complex concepts*, i.e. ones with structure (Margolis and Laurence, 1999: 4-5). Some may take complex concepts to be composed of other, supposedly more fundamental concepts. For example the concept JOEY is composed of the concepts MARSUPIAL and INFANT. Because, on one sense of the word, analysis is decomposition i.e. the breaking down of something into its simpler, more fundamental parts, conceptual analysis is frequently defined as finding out what the fundamental components of our concepts are. In this way, by finding out how our complex concepts are composed, we would presumably

gain a greater understanding of them. It is then easy to see why philosophers<sup>12</sup> have characterised conceptual analysis as this:

CA1: The decomposition of our concepts into their more fundamental components

This characterisation would be inadequate (for our purposes of broadly defining conceptual analysis) if we took it as literal, as it both assumes a restrictive view of conceptual structure and of analysis. Regarding conceptual structure, it rules out *Inferentialism*, which Laurence and Margolis describe thus:

According to this view, one concept is a structured complex of other concepts just in case it stands in a privileged relation to these other concepts, generally, by way of some type of inferential disposition (1999: 5).

Regarding analysis, CA1 taken literally rules out what Michael Beaney calls *transformative analysis*. In describing it he says:

Fundamental to the [Russell's theory of descriptions] is the *rephrasing* of the sentence to be analysed . . . There is nothing decompositional about this type of analysis (2007: 2-3).

In this way, following Quine (1959: 20-1) and Rey (2012), we should view 'decomposition'—and the containment model of concepts it implies—as metaphorical. Even as a metaphor I would say that it is too restrictive (for reasons that I give later on in the essay). A broader and better characterisation of conceptual analysis would be one that presupposes as little about the actual structure of our concepts as possible (or even if they have structure at all!):

CA2: The process of specifying the application conditions for our concepts in a non-trivial way<sup>13</sup>

One of the more recent schools of conceptual analysis that would fall under this

---

12 To take one example: "It seems necessary, then, to regard the world as formed of concepts. These are the only objects of knowledge. They cannot be regarded fundamentally as abstractions either from things or from ideas; since both alike can, if anything is to be true of them, be composed of nothing but concepts. A thing becomes intelligible first when it is analysed into its constituent concepts" (Moore, 1889: 9).

13 I say "in a non-trivial way" because, for instance, the application conditions for the concept BELIEF could be spelled out trivially as: "x is a belief iff x is a belief".

characterisation is known colloquially as the ‘Canberra plan’. It draws primarily from the work of David Lewis and Frank Jackson. (Despite the introduction I just gave to conceptual analysis and also the name itself, Frank Jackson’s treatment of conceptual analysis does not include much discussion about concepts themselves. He instead says “our subject is really the elucidation of the possible situations covered by the *words* we use to ask our questions” (1998: 33). This does not matter for us so much at this point but note that it is odd for the foremost contemporary champion of conceptual analysis to have so little discussion about the nature of concepts.) Jackson characterises conceptual analysis as finding out whether a story told in one vocabulary is made true by one told in another “more fundamental vocabulary” (1998: 28). This characterisation relates to CA1. So that the story told in the “more fundamental vocabulary” would be told with the concepts which supposedly compose the concepts used in the original story. It is also similar to Brandom’s remark that analytic philosophy has at its heart “a concern with semantic relations between what I will call ‘vocabularies’ . . . Its characteristic form of question is whether and in what way one can make sense of the meanings expressed by one kind of locution in terms of the meanings expressed by another kind of locution” (2008: 1). Jackson’s conception of analysis has a much more reductive flavour than this as we shall see. It connects with CA2 in that the application conditions for a concept would be the concept spelled out in the more fundamental vocabulary. Some points that set apart the Canberra plan from other forms of conceptual analysis is (1) its formulations of various philosophical problems and the solutions it offers, and (2) the importance it gives to folk theory<sup>14</sup>. By ‘the folk theory of X’, I mean roughly the ordinary conception of X. Frank Jackson takes it that the folk theory is an “amalgam of individual [conceptions]” (1998: 32n). He states: “To the extent that our intuitions coincide, they reveal our shared theory. To the extent that our intuitions coincide with those of the folk, they reveal the folk theory” (1998: 32).

Concerning (1) the Canberra planner takes many philosophical problems to have a similar form, i.e. a question regarding where (if they are located at all) the so-and-sos are located in such-and-such a description of the world. These are

---

<sup>14</sup> From this point on, when talking about the Canberra plan, I will primarily be talking about Jackson’s view, unless where specified.

called by Jackson *location problems* (1998: 4). It is not immediately obvious where, for example, ethical properties, colours, beliefs, etc. are contained within such a description of the world. The practice of solving these sorts of problems is what Jackson calls *serious-metaphysics*. Jackson's method for solving location problems is by showing how the things-to-be-located are entailed jointly by a definition of that thing and a description of the world in a certain more fundamental vocabulary, one that effectively shows the conditions for the definition being instantiated are met (1998: 4-5). Consider joeys. 'Being a joey' is a different property each from 'being an infant' and 'being a marsupial'. However, we can see that being a joey is nothing over and above the conjunction of these two properties. If there is a true description of the world which specifies infant marsupials then that description entails the existence of joeys, by virtue of our definition of joey (in fact Jackson states that being entailed by an account is a necessary condition for having a place in an account (1998: 5)). In this way we need not posit any separate ontological category for joeys over and above infants and marsupials.

Jackson characterises serious metaphysics as trying to comprehend the world in a limited number of more or less basic notions. By doing this we are not merely drawing up a list of the things that exist and positing a separate ontological category for each. As our example illustrates, we need not posit a separate ontological category for everything that exists because some things exist in virtue of other ontological categories; they are 'nothing over and above' certain other things. A more philosophically stimulating example could be finding out whether the concept BELIEF (a term from our folk psychological vocabulary) has a realizer in the language of (a mature) cognitive science. If our statements about beliefs were not made true by any description in a more fundamental vocabulary then we would have to take beliefs as primitive or be eliminative about beliefs as traditionally conceived (1998: 4).

It is quite easy to see the connection between serious metaphysics and conceptual analysis thus characterised. We characterised conceptual analysis as the process of specifying the application conditions for our concepts in another more fundamental vocabulary. In this sense if we have application conditions for a concept told in a more fundamental vocabulary, we can reduce the concept to

the story told in said vocabulary, so the former is nothing over and above the latter. Let's say the necessary and sufficient conditions for being a joey take this form:

x is a joey iff x is an infant and x is a marsupial

Because the left hand side of the biconditional is truth conditionally equivalent (by definition no less) to the right hand side we can effectively reduce the left hand side to the right hand side. There is no way that something could be a joey without also being an infant and a marsupial. Likewise, if something is an infant marsupial then by definition it is a joey. The realization of a complex concept may be nothing over and above the things specified in the application conditions. In fact the realization of the application conditions would, according to Jackson, entail that the concept is realized. This form of analysis is called *reductive analysis*. It differs from other forms of analysis in that it seeks to reduce certain categories to other more fundamental ones, rather than, for example, show their interconnections (Strawson, 1992: 19-21).

However there are other connections between serious metaphysics and conceptual analysis, hence (2). Jackson sees serious metaphysics as committed to some form of conceptual analysis because in order to perform serious metaphysics we need to 'define our subject'. This defining the subject forms a key part of this style of conceptual analysis, and hence forms the first part of Daniel Nolan's two-steps to Canberra plan analysis (Nolan, 2009):

- (a) We define the subject by gathering our platitudes about the Xs and regimenting them so that we have a neat theoretical role the X plays.
- (b) We look in the world for something that plays this theoretical role; we find out whether our description of the Xs is made true by a description in a more fundamental vocabulary.

So why is defining the subject so important? When we investigate whether Xs exist and in what manner they may exist we are, in a sense, obliged to operate with a certain conception of Xs, i.e. the one that a relevant group operates with. The term Jackson uses for the relevant group is the *folk*. This does not necessarily mean we need to know the man on the street's conception of 'X'; he may not

possess the concept of X to the required degree. Furthermore, the man on the street most likely does not have concepts of neutrons, of rhizomes, or of infinitesimals. These are scientific concepts; the ordinary conception is the one that that people working in the relevant field operate with. So in this sense, when analysing scientific concepts the ‘folk’ are simply those people working in the relevant field, or who are familiar with the literature. Sometimes the folk are the folk in the ordinary sense, as is the case when we analyse the terms of folk psychology. We all partake in the practice of positing mental states; we all mostly follow the rules of folk psychology, so we are all part of the relevant group (1998: 30-31)<sup>15</sup>.

But why are we obliged to operate with the folk conception of Xs? Well if we deviate from the ordinary conception then we have changed the subject. According to Jackson, this is not good at all, because when we ask whether or not beliefs exist we do not ask whether beliefs according to some philosopher’s definition exist. As Jackson states “If I say what *I* mean – never mind what others mean – by ‘belief’ is any information-carrying state that causes subjects to utter sentences like ‘I believe that snow is white’, the existence of beliefs so conceived will be safe from the eliminativists’ arguments”, but then “I have turned interesting philosophical debates into easy exercises in deductions from stipulative definitions together with accepted facts” (ibid).

So how would we define our subject? The way that Jackson and Lewis recommend is to gather our platitudes concerning the Xs and conjoin them into a sentence (Braddon-Mitchell and Nola, 2009: 2-7). This sentence’s job would be to capture our folk theory of the Xs. It would then provide the Xs with a theoretical role, which would make it easier to see if something in the world played said theoretical role and hence could be regarded as an X. Part of defining the subject would consist in reflection on possible cases. Our conjunction of platitudes may align with folk intuitions about actual cases and yet radically differ in how it classifies possible cases. If this were the case with our conjunction of platitudes then it would not have adequately captured our folk theory of the Xs. There are

---

<sup>15</sup> In this chapter I make use of the term ‘folk-concept’. Philosophers usually would just call the former a ‘concept’; however, later on I make use of more unnatural concepts so my use of the former term is just to help distinguish natural concepts from concepts in general.

some issues with this that need to be addressed:

- (i) How would we organise the platitudes into a theory and how do we represent the theoretical role for the Xs?
- (ii) How would we ensure that in both vocabularies we are describing the same thing?

The typical Canberra planner's response to these issues is to appeal to the Ramsey-Carnap-Lewis treatment of theoretical terms. Lewis in his paper *How to Define Theoretical Terms* (1970) seems to only consider scientific theories; however he later applies the same treatment of theoretical terms to folk psychology (1972).

As we have said: conceptual analysis is the very business of finding out whether something in one vocabulary is made true by a description in a more fundamental vocabulary. But how do we decide what vocabulary counts as the fundamental one? Ramsey (or perhaps Lewis's interpretation of Ramsey), as well as helping us define our terms, gives us a suitable framework for doing this. Ramsey introduces the notions of 'primary' and 'secondary' systems. The primary system consists of all the terms and propositions in the universe in question, these constitute "the facts to be explained". The secondary system would be symbols that stand for propositional functions of a theoretical language (these would just be terms and predicates in the theoretical language (Psillos, 1999: 51)). Now<sup>16</sup>, following Lewis, we can interpret the secondary system as consisting of theoretical terms or T-terms, and the primary system as O-terms. Philosophers have interpreted the O-terms in different ways. For example, Carnap takes the O-terms to be observational terms; ones that refer to sense data. Lewis takes the O-terms to be old terms, or original terms; ones that we already understood. My use of O-terms will be closer to Lewis's, with some slight modifications. Lewis applies his method mainly to scientific theories, so T-terms would be theoretical entities such as electrons or quarks. This is why he states that the original terms must have been understood before we introduced the theory. However I am not limiting this method to science. 'Knowledge' may have been a term that was introduced before 'belief' was chronologically, however that does not stop us from analysing knowledge in terms of belief in the way that

---

<sup>16</sup> The following discussion draws heavily from Lewis (1970, 1972).

Ayer does. Another way in which Lewis's treatment of theoretical terms differs from my own is that Lewis views all of the relations between the entities denoted by the O-terms and T-terms as causal. I follow Jackson in not holding to this. In this case the fundamental vocabulary is the one that we can analyse the theoretical terms in, i.e. the one we already understand (the O-terms). For my purposes here, however, it would not suffice to give a definition of a term merely in a vocabulary that we already understand. For a reductive analysis the term must be defined in an ontologically more fundamental vocabulary. I would then add that in order for our Ramsey sentence (I will explain what that is soon) to aid in reduction, we would need the O-terms to be in an ontologically more fundamental vocabulary than the T-terms.

Let's say we want to reduce KNOWLEDGE to a description in a more fundamental vocabulary. First we should gather the platitudes that the folk hold about knowledge.

S knows that P only if they believe P

S knows that P only if P is true

S knows that P only if they are justified in believing that P

Now let's say that we understand the terms 'belief', 'true', and 'justification'. These will then be our O-terms; the ones that we understand already. Our T-term is the term we are attempting to define, so in this case it is 'knowledge'. We can conjoin these platitudes into a theoretical postulate of T, where T is an approximation of the folk theory of knowledge:

T [knowledge is a relation to a proposition P that S stands in when S has a justified true belief that P]

If we replace the T-terms with variables then we have the realization formula of T:

T [x is a relation to a proposition P that S stands in when S has a justified true belief that P]

If the realization formula of T is satisfied by some entities, i.e. if the description

of the variables is satisfied, then we can say that the theory T is realized. If we existentially quantify the realization formula we have the Ramsey sentence of T:

$\exists x T [x \text{ is a relation to a proposition } P \text{ that } S \text{ stands in when } S \text{ has a justified true belief that } P]$

The Ramsey sentence of T effectively says that there is some entity or set of entities that satisfies the realization formula of T. So this Ramsey sentence effectively says there are justified true beliefs. And barring global scepticism this is true, for we have many cases of justified true belief in everyday life.

Now surely (if we can also provide an analysis of justification, truth, and belief that allows for it, and barring a global sceptical hypothesis) under this definition we can say that people have knowledge, for we do have justified true beliefs. However, part of defining the subject (and hence deciding if we are operating with the right definition) is considering whether possible cases of justified true belief would intuitively count as knowledge. As it turns out there are possible cases of justified true belief which do not count as knowledge, i.e. cases of epistemic luck, most notably 'Gettier cases'. Because we would not count these cases of justified true belief as knowledge, by reducing knowledge to justified true belief, we have not adequately defined the subject. If our analysis of a concept is supposed to represent the folk theory of the concept as best as we can, then if there are cases where it conflicts with the folk theory then there is a problem with our analysis (Jackson, 1998: 31-4).

Very plausibly there could be something wrong with our analysis of the folk theory, just as there are problems with the JTB account of knowledge. But another possibility is that there could be a problem with the folk theory itself. For example, there could be platitudes that are inconsistent together<sup>17</sup>, the platitudes just don't describe the world accurately, or perhaps the platitudes only cover everyday situations. In these cases the theory is inadequate. However, does this mean that we should be eliminativists about the concept in question? Not according to Jackson. For him if we define our concept close enough to the folk's conception "to be regarded as a natural extension of it, and which does the

---

<sup>17</sup> Weatherson (2003: 24-5) argues that our platitudes regarding knowledge may be inconsistent.

theoretical job the folk give to [the concept]" (1998: 44-5) then we have successfully defined the subject, regardless of whether we have to shave bits off or add bits on to the theory. In certain cases the theoretical job a concept is supposed to perform can presumably be performed by a number of different variations of that concept.

Jackson draws a parallel between this and Quine's notion of paraphrasing (*ibid.*). For Quine, we paraphrase a sentence of ordinary language by interpreting it into a more logical vocabulary. When we do this we do not aim at synonymy, instead we may, for example, aim at simplicity, or the reduction of ambiguity. Quine states that although the sentence and its paraphrase may be used synonymously, that isn't to say that they are synonymous; the most we can say here is that both sentences perform the same purpose for the speaker (1960: §46). Now the connection with Jackson's account is that in interpreting the folk theory we may not necessarily aim at synonymy either. Our folk concepts play some theoretical role, to use Jackson's example, our concept of personal identity plays the role of governing our personal relations, our social institutions of reward and punishment, etc. Although our folk conception of personal identity may be Dualist in nature, it need not be in order for it to prevent questions like this from arising: "Why should I pay that debt from five years ago? I was a completely different person back then!" So, to run with Jackson's example, if we want to analyse our concept of personal identity we should try and stay faithful to the folk theory of it, but also we should look at the role that the concept plays in the folk theory. If that theory turns out to be false, or even if it clashes with our prior commitments (in this case physicalism), we may still be able to show that something can occupy the theoretical role (1998: 44-5).

This picture of conceptual analysis is inadequate as I shall show in the next section. I take the Canberra plan—as exemplified by the work of Jackson and by Nolan's two-step—to be primarily descriptive. However, Canberra planners may be much more receptive to conceptual revision than I suggest. In that case, I would say I am arguing against a particular kind of descriptive analysis. This, I think, would be an important view in the vicinity. Nevertheless, I think there is much textual support for taking Canberra planners to be engaging in a heavily descriptive exercise. For instance, Jackson places a lot of emphasis on

our intuitions about possible cases. He says of William Lycan's account of beliefs:

I of course hold against Lycan that if we give up too many of the properties common sense associates with belief as represented by the folk theory of belief, we do indeed change the subject, and are no longer talking about belief. The role of the intuitions about possible cases so distinctive of conceptual analysis is precisely to make explicit our implicit folk theory and, in particular, to make explicit which properties are really central to some state's being correctly described as a belief. (1998: 38)

This is good evidence for Jackson taking conceptual analysis to be descriptive in nature. Sure, he allows some room for interpretation, but it seems that, for Jackson, one cannot move away from what is central to the folk concept.

## Problems for conceptual analysis

In this section I will outline a problem for Jackson's version of conceptual analysis, from Timothy Williamson. Williamson casts doubt on the prospects for finding an exact definition by way of an inductive argument; philosophers have been trying to provide analyses of concepts for years with no real success, so why should we trust this method at all?

### *Williamson and the inductive argument*

Williamson argues that we have no reason to accept that we can provide a non-trivial analysis of most concepts. Williamson's argument is primarily directed towards analyses of KNOWLEDGE. However he generalises his argument to

other concepts: “Attempts to analyse the concepts *means* and *causes*, for example, have been no more successful than attempts to analyse the concept *knows*, succumbing to the same pattern of counterexamples and epicycles” (2000: 31). What constitutes a non-trivial<sup>18</sup> analysis for Williamson is a conjunction of other concepts that somehow jointly specify necessary and sufficient conditions. Williamson does also talk about conceptual identity as a criterion for the correct analysis of a concept. For instance, he says that KNOWLEDGE is a mental concept, and so a non-mental<sup>19</sup> concept like JUSTIFIED TRUE BELIEF—which involves non-mental constituents—would necessarily be distinct from it. This aspect of Williamson's argument I will not address, partly because—jumping ahead slightly—my solution to the inductive argument relies on philosophical analysis not being bound by the analysis being identical to the concept analysed. Furthermore, stating correct application conditions for a concept, regardless of conceptual identity, would at the very least be an interesting. Williamson's argument that one cannot find the application conditions for the concept of knowledge partly comes from his thought that one cannot find an identical non-mental concept:

The present account does not strictly entail that no analysis of the traditional kind provides correct necessary and sufficient conditions for knowing. But once we accept that the concept *knows* is not a complex concept of the kind traditionally envisaged, what reason have we to expect any such complex concept even to provide necessary and sufficient conditions for knowing? (2000: 30).

In this way we can merely take this criterion as support for the inductive argument, rather than being essential to it. So henceforth I will take Williamson's point to be that we cannot articulate necessary and sufficient conditions for most concepts, and I will not put much emphasis on conceptual identity<sup>20</sup>.

---

18 Though not necessarily a correct analysis.

19 So he argues (2000: 27-31)

20 Cassam, identifies two strands of argument in the relevant passages from Williamson alongside that already stated. One is the distinct concepts argument; another is the false expectations argument. The first concerns concept identity, the second concerns us not expecting an analysis of knowledge because no other interesting philosophical concepts has been analysed in the way we want. These seem quite restricted to Williamson's treatment, of knowledge, so they won't be directly relevant to my discussion of concepts in general. However I will assimilate the false expectations argument with the inductive argument (2009: 12).

One part of Williamson's argument for this point is based on the history of conceptual analysis being a history of failure. Usually whenever someone offers an analysis of a concept, a counterexample to that analysis is produced. And this supposedly has been the case in epistemology for some time<sup>21</sup>. Ever since Gettier proposed a counterexample to the 'justified true belief' account of knowledge ever more complex analyses of knowledge have appeared with ever more complex counterexamples soon following. Williamson inductively reasons that because we have not so far been able to successfully analyse knowledge in terms of linguistic descriptions, we would have little reason to think that it would be possible to do so. Williamson says that this is not only the case in epistemology but also throughout philosophy: "Not all concepts have such analyses [in terms of more basic notions] . . . the history of analytic philosophy suggests that those of most philosophical interest do not" (2000: 31).

Of course Williamson's evidence may only show that it is very hard to analyse a concept and that the road to an adequate analysis of a concept is long, and paved with counterexamples. But without an example of such an analysis we would not have much evidence at all to support this suggestion.

Jackson has a response to Williamson's argument. He says that the world is huge, complex and we make sense of it through similarity regions. Conceptual analysis is trying to capture the same similarity region in two different ways, e.g. the region picked out by 'joey' is the same as the one picked out by 'infant marsupials'. These patterns would constitute the application conditions for JOEY. Jackson holds that it should be no surprise that conceptual analysis is hard because there is no reason to think "it should it be obvious that different sets of ingredients put together in appropriately different ways carve out the same region" (2005: 132). Of course this shows why conceptual analysis is hard; however, it does not give us any reason to think that it is possible. What the conceptual analyst would ideally want in a response to Williamson is reassurance that their method is worth pursuing. If Jackson is right in saying that conceptual analysis is hard, then perhaps it would not be too much of a stretch

---

21 I myself think that the success of the counterexamples is overstated. Sure, potential counterexamples have been offered for every well-known analysis of knowledge but as the analyses become more refined the counterexamples get more outlandish.

for it to be *too hard*; that is, too hard for us to ever reach a successful analysis of a concept.

However, another part of Jackson's response to Williamson is intriguing, and I believe contains the seed to a response to Williamson. Jackson holds that the reason there seems to be so much disagreement over what counts as knowledge (i.e. why it seems like we cannot have a successful analysis because of counterexamples) is that there is not a single fixed pattern across the space of possibilities that picks out 'knowledge'. Instead there are many candidate patterns: Justified true belief, process reliabilism, Nozickian sensitivity, etc. and "for each candidate concept there is an analysis" (2005: 135). In most cases it does not matter what one of these we operate with; they all give the same results in everyday life. This is how Jackson allows for an analysis of KNOWLEDGE. His response is to say that we can analyse a concept like knowledge; different philosophers have different concepts that they call knowledge, and they can all provide an analysis of knowledge according to their own concept.

This response is intriguing because of his emphasis on defining the subject. Jackson's style of doing conceptual analysis, as I interpret it, seems to primarily aim at capturing the folk theory. When we try and capture the folk theory we do not want to analyse some philosopher's concept of knowledge; we want to analyse knowledge insofar as individual's conceptions of knowledge overlap (1998: 31-3). If there is only a significant overlap over our classification of everyday cases then I assume Jackson would identify that overlap with the folk theory: "to the extent that our intuitions coincide, they reveal our folk theory" (ibid). That way, as the different philosopher's analyses are co-extensive with regard to everyday cases, the different philosophers would just be filling in the blanks of the folk theory. In this way their analyses can be seen as an extension of the folk theory.

However, although Jackson may be right in saying that different philosophers have different conceptions of what counts as knowledge, this does not entail that they each have different concepts of knowledge. Two people may have different opinions on what counts as an X without operating with two different concepts of X. Consider Williamson's example: "someone who relies on

The Guardian in categorising politicians as ‘trustworthy’ or ‘untrustworthy’ is categorising in a different way from someone who relies on The Daily Telegraph, but it does not follow that they mean different things by the word ‘trustworthy’” (2009: 130). This is equally applicable in the philosophy room: When epistemologists get around the table and throw counterexamples at one another they do not respond by commenting “well that is not a counterexample to *my* analysis, because for me knowledge is just XYZ”. When epistemologists are offered such counterexamples they feel obliged to address them in some other way than just biting the bullet and saying that, for example, Gettier cases are actually cases of knowledge. There seems to be some kind of shared concept we are aiming at (I later examine where the power behind these counterexamples comes from).

Jackson’s response should also show why counterexamples to certain proposed analyses of knowledge seem so decisive. For example justified true belief does not seem as good as the less fluky true belief reached by a reliable method, no doubt because of the power of Gettier cases. He gives us two (non-mutually exclusive) possibilities of where the power of counterexamples comes from:

- (1) Counterexamples show that our analysis is flawed because it shows that our folk concept is not captured by the analysis
- (2) Counterexamples Show that our analysis is flawed because although we may have succeeded in analysing the concept we operate with (this may or may not be the concept most of us operate with); the concept we operate with is inferior to another concept that is useful for the purposes we want the original concept for.

We can call instances of the first *type-1 counterexamples* and instances of the second *type-2 counterexamples*. In the first instance his response to Williamson may not work. If his purpose is to capture the folk theory of a concept, and he is presented with good type-1 counterexamples, then it seems that he has not adequately captured the folk theory; there would be more to the folk theory than the everyday cases, otherwise where would the counterexamples of type-1 come from? In the second instance I would say that he escapes Williamson’s criticism;

but only if there were no good type-1 counterexamples. If his aim is to capture the folk theory and there is a counterexample of the type-2 variety, then we have shown a problem with the folk theory, not with his analysis.

However, although the Gettier cases are certainly type-2 counterexamples to the JTB account of knowledge (as they show theoretical defects of JTB), there are good reasons to think they are type-1 counterexamples too. I take it that type-1 counterexamples take their power from our intuitions about possible cases. This is because the type-1 counterexamples would come from our understanding of the concept at hand. Intuitions also seem to be where the Gettier cases get their potency, so this is one feature Gettier cases would share with paradigm type-1s. Of course, type-2 counterexamples may also get some of their strength from intuitions, for instance we might have intuitions of theoretical virtues such as simplicity, exactness, etc. Nevertheless, I would say Gettier cases are type-1s because of the way we invoke the intuitions. We do not generally say “The Gettier cases show that the JTB account of knowledge is flawed because they are cases which do not feature the non-flukiness which we would want knowledge to have” (although we could); we instead say something along the lines of “The Gettier cases show that the JTB account of knowledge is flawed, because in them people have justified true beliefs without having knowledge”. This at least muddies the water, so that it is entirely possible—in fact likely—that the counterexamples to the JTB theory have been type-1s, and so there would be a problem with the analysis, and not the folk theory.

### *Conceptual analysis as a verbal enterprise*

Another objection to conceptual analysis is that engaging in it is a merely verbal pursuit, so is at worst pointless. (Some may insist I use the word 'conceptual' instead of 'verbal' here as we are talking about conceptual analysis, but both can effectively do the same work in this objection. Certainly in some cases,

conceptual differences may even be what lead people to verbally dispute.) Recall that in the last chapter, we endorsed Chalmers's definition, or rather a variant thereof, as our definition of a verbal dispute. However, Chalmers holds that his view of verbal disputes is somewhat deflationary about Conceptual analysis. He states:

This picture leads to a certain deflationism about the role of conceptual analysis (whether a priori or a posteriori) and about the interest of questions such as “What is X?” or “What is it to be X ?” Some component of these questions is inevitably verbal, and the nonverbal residue can be found without using ‘X’ (2011: ).

Conceptual analysis is at its most basic, finding the application conditions for our concepts. An example of an expression is found in this remark from Frank Jackson which I quoted earlier: “conceptual analysis is the very business of addressing when and whether a story told in one vocabulary is made true by one told in some allegedly more fundamental vocabulary” (1998: 28)<sup>22</sup>. The verbal element concerns the application conditions which we attach to a concept of philosophical interest. There are often competing sets of application conditions in the vicinity, and the choice between these, the argument goes, would be a verbal (or conceptual) one. Take one of Chalmers's examples:

One frequently finds verbal elements in disputes over the formulation of physicalism. For example, some physicalists hold that physicalism is the thesis that everything supervenes on the properties invoked by a completed physics (whether or not they are mental), while others hold that physicalism is the thesis that everything supervenes on the properties invoked by a completed physics and that these properties are nonmental (2011: 533)

The verbal element in formulating physicalism then, is the question of what the application conditions for PHYSICALISM (or PHYSICAL) are. But then the merely verbal element *is* the conceptual analysis, or at least a large part of it. As we saw in the last chapter, verbal disputes are nonsubstantive and engaging in them is a pointless activity. This, the possible objection goes, should lead to deflationism about conceptual analysis.

---

<sup>22</sup> These two characterisations are compatible. Whether one a story in one vocabulary is made true by another will depend upon the application conditions of the terms in the first story.

As an aside, this is perhaps not the precise view of Chalmers, for he *does* see a place for conceptual analysis.

The model is not completely deflationary about conceptual analysis. On this model, the analysis of words and the associated concepts is relatively unimportant in understanding a first-order domain. But it is still interesting and important to analyze conceptual spaces: the spaces of concepts (and of the entities they pick out) that are relevant to a domain, determining which concepts can play which roles, what the relevant dimensions of variation are, and so on (2011: 539).

But he is clear about what he sees as the limits of conceptual analysis:

On the picture I favor, instead of asking “What is X ?,” one should focus on the roles one wants X to play and see what can play that role. The roles in question here may in principle be properties of all sorts: so one focuses on the properties one wants X to have and figures out what has those properties (2011: 538).

[The view expressed in the quote before last] leads naturally to a sort of conceptual pluralism: there are multiple interesting concepts (corresponding to multiple interesting roles) in the vicinity of philosophical terms such as ‘semantic’, ‘justified’, ‘free’, and not much of substance depends on which one goes with the term (2011: 529).

In this way, the question of what the particular application conditions are of a concept is relatively unimportant, compared to the question of what sets of application conditions are in the vicinity and what roles they can play. However, not everyone who accepts the idea that conceptual analysis is pointless in the same way as verbal disputes are would accept Chalmers's conceptual pluralism, so I will not include these qualification of Chalmers's account in the main objection of this section. The objection put to the Canberra planner then is that conceptual analysis has the same feature that makes verbal disputes verbal, and hence it may suffer from the same type of pointlessness.

The charge of verbalness may have more specific negative consequences for the Canberra plan. Consider the emphasis that the Canberra plan style puts on folk-theory. As stated in the section “Conceptual analysis and the Canberra Plan”, what the Canberra planner (ideally) does is gather our various platitudes regarding some term 'X', combine them so we have a theoretical role for X, and

then sees if there is anything that plays this theoretical role. One way the objection affects this style of analysis regards the various ways we can interpret the folk-theory. As our folk theory of X mostly regards everyday situations, there may be considerable room for interpretation in our analysis. Jackson states: “What compatibilist arguments show, or so it seems to me, is not that free action as understood by the folk is compatible with determinism, but that free action on a conception near enough to the folk's to be regarded as a natural extension of it, and which does the theoretical job we folk give the concept of free action” (1998: 44-5). The way an analysis would be a 'natural extension' of the analysed concept—as opposed to an unnatural extension of it—would, I think, be in its categorising everyday examples in accordance with the ordinary conception. So perhaps there are many natural extensions of a particular folk-concept, and the choice between these various analyses will be merely verbal.

## Analysis and decomposition

One thing all of these criticisms have in common is a certain view of analysis. They all view analysis as aiming to capture a folk theory through our application conditions for it.

This I think expresses the *decompositional* view, which entails analysis as the breaking down of a complex concept into its 'proper parts'. These proper parts are would be other concepts. As stated in the section “Conceptual analysis and the Canberra Plan”, decomposition can be taken as metaphorical: In this way a complex concept would 'contain' other concepts in the sense that other concepts would have to be instantiated in order for the complex concept to be instantiated. Our concepts 'containing' other concepts would be nothing over and above the other concepts featuring in their application conditions. We may think that JOEY 'contains' MARSUPIAL and INFANT, as one cannot correctly apply JOEY across the board without being able to correctly apply MARSUPIAL and INFANT as well, as JOEY would be the intersection of MARSUPIAL and

INFANT. So, if we correctly apply MARSUPIAL and INFANT to the same thing then that thing is a joey.

What is essential to decomposition is that we start off with a more or less exact concept to analyse, and that we articulate necessary and sufficient conditions for its use. Even Jackson seems to have something close to this in mind because of his insistence on capturing the folk theory, or something suitably close to it. Notice that on his way of admitting that there can be many analyses of KNOWLEDGE, he still has to identify each concept of knowledge with a particular person's 'theory' of knowledge.

Decomposition does not seem to exhaust the conceptions of analysis. It seems that certain things can be analysed in various ways, take Michael Dummett's example of a geographer carving up a country into different regions: "One and the same country might be subdivided into regions in different ways for different purposes—in terms of geological structure, the kind of terrain, types of vegetation or climate, the languages, religions or cultures of the inhabitants, etc." (1981: 263). This example does not really express the specific kind of analysis I outline and defend however. In this example we already have all the information there ready to be decomposed, it is just a matter of how we choose to do it. (Also in this example we are not analysing a concept, rather a concrete object.) The conception of analysis I will defend later in the essay has a more ampliative feel.

## Counterexamples

Counterexamples are important to Williamson's criticism. Recall Jackson's distinction between those counterexamples that show we have not captured the folk theory (type-1), and those that show there is a superior concept to that expressed by our analysis (type-2). When an analysis is faced with a counterexample it would be hard to tell which of these categories, it falls into<sup>23</sup>. I

---

23 It certainly seems plausible that a counterexample could fall into both categories.

am quite sure that there are examples of both types. An example of the first type is the Gettier case. It seems to be a member of the first type because of the widespread agreement that it is a genuine counterexample to the JTB account of knowledge. This counterexample arose from testing a definition against our intuitions about a possible case. Gettier cases count as counterexamples to the JTB account of knowledge because they show situations where a proposed definition and our intuitions disagree. I hold, as Jackson does, that our intuitions about possible cases are a good guide to our folk theory (1998: 31-4). Hence, if we have very strong intuitions about Gettier cases counting as knowledge then JTB does not accord with our folk theory. The first type of counterexample may even connect with the second in certain cases; the counterexamples may come from the same source. It may be that our folk-theory of knowledge does not have determinate application conditions written into it; rather it aims to exclude cases of epistemic luck for reasons of knowledge being reliable. So a situation which features epistemic luck may count as a type-1 counterexample because it conflicts with the folk theory's aim, i.e. it involves epistemic luck; but it may also be a type-2 counterexample because the folk theory's aim may be to exclude cases of epistemic luck without explicitly identifying what type of epistemic luck is involved. In this case the folk theory may be dynamic; it may change depending on the cases of epistemic luck it becomes aware of. However it would retain the same aim, i.e. excluding cases of epistemic luck.

I interpret Williamson as focusing on type-1 counterexamples. This is because they seem to be the only ones that are decisive in showing that a given analysis is false (if of course our primary purpose is to spell out application conditions for the folk theory). If we are operating with our folk concept of knowledge, and it is part of that concept that Gettier cases do not count as cases of knowledge then the conditional 'if K is knowledge then K is not a Gettier case' would presumably be a priori; we could know it to be true by possession of the folk concept (Chalmers, 2012: 13).

If we are engaging in decompositional analysis, then type-1 counterexamples would be the main ones we would have to watch out for. However, if the purpose of our analysis is not to do this—if it is *ampliative*—then these counterexamples will not seem so powerful (what exactly the purpose of

our analysis is we will look at in the next section). If we see the folk theory as merely giving our analysis a starting point then large deviation from the folk theory may be acceptable. As stated, the primary purpose of Jackson's account seems to be spelling out or approximating the application conditions for our folk concepts as they are. He states: "[for example] what we are seeking to address is . . . whether intentional states *according to our ordinary conception*, or something suitably close to our ordinary conception, will survive what cognitive science reveals about the operations of our brains" (1998: 31). So a more ampliative view of analysis may allow for our analysis of a concept to be quite far away from the ordinary conception of it, something that I doubt Jackson would allow for. What constitutes something being suitably close to or unsuitably far from our ordinary conception according to Jackson is a vague matter. However, recall that Jackson says this: "What compatibilist arguments show, or so it seems to me, is not that free action as understood by the folk is compatible with determinism, but that free action on a conception near enough to the folk's to be regarded as a natural extension of it, and which does the theoretical job we folk give the concept of free action" (1998: 44-5). This is permissible for Jackson as compatibilists show that "a compatibilist substitute does all we legitimately require of a concept of free action". So perhaps for Jackson changing the subject is permissible when the analysis can be seen as a natural extension of the folk concept, and it does the theoretical work that the folk wish the concept to do.

I am not so sure that this would allow Jackson to make significantly large revisions to a folk theory. He states that "it is, to one extent or another, vague as to which concept we use the word 'knowledge' for, and in practice it often does not matter—that which we are certain of is very often that which we arrive at by a reliable process and which is not true by accident" (2005: 135). These various concepts all give similar results in everyday life with regards to our application of the term 'knowledge'; in this way they are all near-realizers of the folk concept; they fill in the theoretical space that the folk theory leaves empty. A far realiser would perhaps be one that differs from the folk theory in classifying everyday cases. An example of conceptual revision (that I talk about in the next chapter) that replaces a concept with a far realiser is perhaps Carnap's example concept FISH (which applies to all and only water dwelling animals) being replaced by

the concept PISCES (which excludes whales, sea snails, etc.). This does not seem to me to be a natural extension, nor a natural limitation of the concept; it seems to be an almost complete replacement. In the next section I shall outline a kind of analysis that allows for large amounts of revision

## Carnap and explication

Explication, in Carnap's terms, is the "task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept". The former concept is called the *explicandum* and the latter is called the *explicatum*. The explicatum must have explicit rules for its use so we know when it is correctly applied or not<sup>24</sup> (1950: 1).

An historical example that Carnap gives of explication is the taxonomy of sea creatures (1950: 5-7). It used to be that the term 'fish' applied to water dwelling creatures. However, (according to Carnap's story) later taxonomists replaced the concept FISH, which applies to all water-dwellers, with the more exclusive concept PISCES (this is the term Carnap uses for the revised fish concept). The extension of PISCES, unlike that of FISH, does not include whales, dolphins, otters, clams, sea-slugs, jelly-fish (despite their primitive name), and manatees. It is instead limited to water dwelling animals with a certain shared ancestry; perhaps, though Carnap may not have thought this, PISCES is a natural kind concept. The problem with FISH is that it is not particularly fruitful, for example when it comes to making scientific predictions. PISCES beats FISH in this regard. As we can make good predictions, e.g. if something is a pisces then it will be a vertebrate, it will live in water, it will have such-and-such genetics, etc.

---

24 An explication does not need to significantly revise the target concept. We could have an explicatum that specifies application conditions for the explicandum without deviating from it at all. In this way it would be extensionally identical to a traditional analysis of a concept. An explicatum is distinguished from traditional analyses by its purpose, i.e. satisfying Carnap's four criteria, rather than capturing the folk theory. It may well capture the folk theory, but we would have little reason to expect that any analysis would, for reason already stated.

because it is largely defined by these properties. The only reliable prediction we can make with the concept FISH is that its members will live in the water. The criteria for a good explicatum are (1) that it be similar to the explicandum, (2) that it be exact, (3) that it be fruitful, i.e. useful for making general statements about the members of the kind in question, and (4) that it be as simple as possible (ibid).

(As we have seen, the advantages that PISCES has over FISH fall under the fruitfulness category, however a problem arises here concerning the exactness of FISH. Carnap states that in “a problem of explication the datum, viz., the explicandum, is not given in exact terms; if it were, no explication would be needed” (1950; 3-4). However FISH does seem to be an exact concept as the corresponding bi-conditional shows:

x is a fish iff x is a water-dwelling animal<sup>25</sup>

If FISH indeed is an exact concept in this way then there is a way we can deal with this. We can simply say that the explicandum in this case FISH is exact however there is a concept that satisfies the other criteria more adequately i.e. PISCES’s satisfaction of fruitfulness.) An explanation is needed here of the four criteria.

### *Similarity*

We can cash out similarity to the explicandum in a few different ways.

*Extension:* Similarity between two concepts could be cashed out in terms of the things they applied to. If the two concepts have a significantly overlapping extension then they would be similar. FISH and PISCES have an overlapping extension however FISH would apply to many more

---

<sup>25</sup> This concept may not be exact, as there are some animals like hippos and penguins that spend long periods of time inside and outside of the water. Hence it may be vague as to whether they are in the extension of FISH or not. However the solution I outline to this problem above I think would apply to any exact but imperfect folk concepts (like FISH).

animals. In fact the extension of PISCES would make up only a small percentage of the extension of FISH (in the actual world).

*Purpose:* Similarity could be cashed out in terms of the theoretical work we want the concept to do. For example, FISH and PISCES are similar in this way because both are used for the taxonomy of water dwelling creatures. The difference is that PISCES is restricted to a particular type of water dwelling creature. This still seems like these two concepts have a bigger similarity in purpose than if one was for taxonomising land dwelling creatures.

*Definition:* Similarity could be cashed out in terms of the application conditions the two concepts share. For example BITCH is similar to VIXEN in that both share the necessary conditions 'being female' and 'being a canine'. However the difference is that VIXEN also builds in the condition 'being a fox'. FISH and PISCES are not definitionally similar at all, apart from water dwelling serving to fix the referents of PISCES.

### *Exactness*

Carnap states that the rules for a explicatum's use must be exact. I take him to mean that the rules for the use of the explicatum must be consistent and must have a more determinate extension over possible situations than the explicandum. This way we would be able to say with relative certainty whether or not the concept is correctly applied. There should, for example, be very few obscure cases where it is indeterminate whether the concept can be applied or not.

### *Fruitfulness*

The fruitfulness here Carnap says consists in the concept being useful for making many universal statements. This needs expanding. The examples of fruitful universal generalizations that Carnap uses are scientific laws and logical theorems. Take the made up concept SKEG. The criteria for falling under this concept are that one be a human who is both ten feet tall and five feet tall at the same time. We can make lots of universal statements using SKEG, e.g. 'no chimpanzee is a SKEG', 'no human is a SKEG', 'all skegs are human', etc. However, these universal statements are not useful in any way. So a further qualification must be that the concept be useful for making *good* universal generalizations.

What Carnap has in mind when he talks about fruitfulness seems to be related to logical or empirical sciences; he does not seem to include any considerations on what would constitute fruitfulness for a philosophical concept (no doubt this is because of his overall project). I would say that in the case of philosophical concepts the most fruitful concepts will be one ones that give us solutions to location problems. I will elaborate on this later on in the chapter.

### *Simplicity*

Carnap does not expand on the notion of simplicity. I think it in part is that our definition should be as short as possible, unnecessary clauses would not be helpful at all. I also take the simplicity of a concept to partly consist in its *naturalness*. Which means that we should not in our definition include gerrymandered properties. These should generally be avoided, for example, we could define a joey as an infant marsupial, but we could also get the same intension by defining it as a 'schminfant' (infant and (either infant or adult but not both)) 'schmarsupial' (marsupial and (either marsupial or feline but not both)). This gives us the same class of creatures, but does it through dubious properties, which are less fundamental than 'infant' and 'marsupial' because of their reliance on disjunctions with other properties like 'feline'.

The more fundamental the properties we use, the more effective our ontological reduction. Part of the rationale for ontological reduction is the project of being able to have a class of entities on which the existence of all other entities supervenes. For Jackson this would be the physical vocabulary. If we define our concepts with using gerrymandered properties we are just adding in steps it would take to effectively reduce the concept to, for example, the language of physics.

### *Explication and the Canberra plan*

Before we go on to explicitly reply to Williamson's argument against conceptual analysis it will be helpful to show how the Canberra plan and Carnapian explication complement and differ from each other. Recall that Jackson holds that many problems in philosophy can be described as location problems, the problem of where the Xs are. The process of solving these problems is a form of ontological reduction called serious metaphysics. I largely agree with Jackson in this. However, as we have seen, the analyses his largely descriptive framework provides are susceptible to Williamson's objection. I shall wait until the next section to explicitly show how explication gets over these problems, but for now I will say getting over these problems is an advantage the Canberra planner gets from adopting a form of explication. The following are two ways that I think show how Carnapian explication and the Canberra plan can mesh with each other.

*Explication makes solving location problems easier.* Carnap himself was famously opposed to metaphysics, and he most likely introduced explication with his overall project in mind. However, this is not to say that explication could not be separated from Carnap's overall project. In fact, Carnapian explication can be very useful for what Jackson calls serious metaphysics i.e. conceiving of the world in a limited number of ingredients. Consider Jackson's entry by entailment thesis. This states that the existence of certain things is entailed by an account of

the world. If we define a joey as an infant marsupial then the existence of an infant marsupial entails the existence of a joey. Whether or not the existence of X is entailed by an account of the world is dependent on the definition we have for X. As explication allows us a lot of leeway in constructing a definition of X, we would also have more leeway in showing where Xs are located. In this way there may be various ways of solving philosophical problems. A reductive explication would simply be an explication of some concept in an ontologically more fundamental vocabulary.

It is not necessary for an explicatum that it be readily reducible. As Carnap states: “The only essential requirement [for an explicatum] is that the explicatum be more precise than the explicandum” (1963: 936). However, as the possibility of solving location problems is one of the main advantages of explication, an explicatum that was not reducible would seem to not go far enough. An example could be the stating of exact application conditions for a concept where the application conditions make reference to mystical properties.

*Reflection on possible cases could help us explicate concepts:* For Jackson, reflection on possible cases is part of defining our subject. If we find a possible situation where our analysis produces unintuitive results then our analysis is defective. As we have shown, an explication does aim at producing a new concept that is similar to the explicandum, so reflection on possible cases would serve the purpose of elucidating the folk theory in both the Canberra plan and in explication. However, as we have seen, an explicatum must also satisfy other criteria. Reflection on possible cases could also serve an important purpose by refining these other criteria. It might be that an explicatum may seem fruitful when we consider actual cases; however it may be that in possible cases the definition is not fruitful. We may, for instance, want a concept of personal identity to govern our personal relations, our social institutions of reward and punishment, our debts that we incurred many years ago, etc. But there may be a rather large collection of possible situations where our definition of personal identity does not do this. In this way it may serve us well to revise our definition.

To illustrate all this we can posit a revised version of Nolan’s Canberra ‘two-step’ that explicitly takes inspiration from Carnapian explication:

- (a) We define the subject by gathering our platitudes about the Xs and using them to elucidate our folk concept (the explicandum) of the Xs. If the concept is inadequate, then we would posit a replacement concept which adequately satisfies Carnap's four criteria (the explicatum).
- (b) We look in the world for something that satisfies the role of the explicatum of X; we find out whether our description of the Xs is made true by a description in a more fundamental vocabulary.

Call this the Carnapian two-step<sup>26</sup>. As we can see, the main difference between this and the original version is in the detail included in the first step. Whilst the first is less specific when it comes to the regimentation of our platitudes into a theoretical role, the second version has clear criteria of what counts as an acceptable analysis. The second version seems to be more liberal with what may count as an analysis of X because of it being able to go beyond near-realiser; however the Carnapian analysis would presumably be more definitely formulated.

One difference between the Carnapian two-step and the original is in the way they respectively treat counterexamples. A few sections ago we drew a distinction between those counterexamples which are situations where an analysis is inadequate because it is not a strict analysis of the concept as we employ it (type-1s) and those which count as counterexamples because the analysis fails to meet other criteria related the purpose of our analysis (type-2). Both the Canberra and Carnapian two-step are open to type-1s, for they both require a degree of similarity to the target concept, but they carry less weight for the Carnapian version. This is because the criteria other than similarity to the explicandum may act as counterweights for the Carnapian two-step; if an explicatum fulfils the other criteria (fruitfulness, exactness and simplicity) then it may still be a good analysis. Observe how the two methods would differently treat Dennett's account of intentional states. A type-1 counterexample to this theory, for instance, thermostats believing the room is too hot , would be a lot more serious if, like the Canberra planner, we were trying to keep close to our folk

---

<sup>26</sup> Though, Carnapian explication should not be restricted to this two step. This two-step is essentially just Carnapian explication with answering location problems being part of the criteria for a fruitful and simple analysis.

concept. However, if we were trying to explicate intentionality, then the unintuitiveness of Dennett's account would be weighed up against its simplicity, exactness, and fruitfulness. Both methods would take seriously type-2 counterexamples, such as when the explicatum seems simple but which we would have to posit epicycles for it to work across certain cases. An example could be that Dennett's analysis of what constitutes a true believer may work just fine across everyday cases, however, in the case of strangely behaving aliens we would have to identify their beliefs with erratic patterns of behaviour. We may not be able to predict the way they will act simply by taking the intentional stance to them; it would require other work as well. However, because the Canberra Planner's primary aim is in giving an analysis of our folk-concept, then they would not take this counterexample as seriously as would the Carnapian.

In essence, the reason why Carnapian explication is not as susceptible to counterexamples as Jackson's method of analysis is because of their different purposes. Capturing the folk theory is less of a flexible goal than satisfying Carnap's four criteria. This is because of the simple reason that if an explicatum does not satisfy one of these criteria, then it may still be adequate because of its satisfaction of the other three criteria. Furthermore, there may be numerous analyses we could choose from, which all satisfy Carnap's four criteria to differing extents. However, if our purpose is primarily to capture the folk theory then if we do not do that our analysis is unsuccessful. There is also a sense in which capturing the folk theory is a much harder task than constructing a concept that, for example, can easily be reduced to more fundamental concepts (not merely because philosophers have been trying to capture folk concepts for years to no avail). The ordinary conception of something is spread across many different minds, across a large area, may not directly investigated by someone, and would most likely change gradually with time. Revising a concept so it can be reduced to something more fundamental, however, seems much easier if we are allowed to deviate from the folk conception. One could practically do this from the armchair!

*Williamson revisited*

Williamson argues that we should have no reason to believe that we will ever get a successful analysis of a philosophically interesting concept, because we have been trying for years to no avail.

Of course in order to assess Williamson's argument we would have to see what would constitute a successful analysis. It seems that for Williamson a successful analysis would be one that specifies the necessary and sufficient conditions for the application of a concept as it is; one which is not open to type-1 counterexamples.

I agree with Williamson in a sense: I do not think that we have succeeded in decomposing many (if any) interesting philosophical concepts in this manner. However, this is not our only criterion for a successful analysis. Recall, that type-1 counterexamples are ones showing that our analysis does not gel with our folk theory. If our analysis does not primarily aim to capture the folk theory then these kinds of counterexamples will not be fatal to our analysis. One consequence of the Carnapian picture is that there is no one correct explication of a concept. A successful explication is one where we replace an inexact concept with an exact one that meets our four criteria to a satisfactory degree. There may be various ways in which we can do this; some of which are better for different purposes than others. And if we interpret previous efforts to analyse philosophical concepts as explications, no doubt, some of them will be successful in this regard. Some may be lacking in certain areas but do so well in other areas that they may be regarded as successful. Consider the amount of analyses of concepts that have been largely discarded because of intuitions about possible cases. Many of these may actually be successful because of the way they satisfy the other criteria. In this way we cannot really say that the history of conceptual analysis is a history of failure; it has perhaps only failed to produce decompositions of concepts, not analyses in the wide sense.

## *The deflationary argument revisited*

The deflationary argument says that the choice between different analyses of a concept is a merely verbal one. So to argue over analyses would be pointless in the same way as a verbal dispute would be. Let's analyse this claim. Recall the definition of verbal disputes we endorsed in the last chapter:

A dispute over S is broadly verbal when, for some expression T in S, the parties use (or are disposed to use) T in different ways, and the dispute over S arises wholly in virtue of this divergence in the usage of (or the disposition to use) T (2011: 522).

The verbalness of the dispute comes from the way in which it arises; specifically, from linguistic divergence between the disputants. That mere verbal disputes arise in this way is also the source of their nonsubstantiveness, as usually the disputants will be trying to answer a first-order question and mere linguistic divergence would not appropriate for grounding this type of dispute.

This point is relevant because it leads to this question: In virtue of what could a dispute between two conceptual analysts be substantive? In a merely verbal dispute, the dispute is nonsubstantive because of the disputants believing there is something deeper to their dispute than a mere choice of words. In this way, the charge against the conceptual analysts would perhaps be that they are not self-conscious about what would ground possible disputes they would have. They would think that there is something more to the question of the application conditions for a concept X, than the mere use of words. So perhaps conceptual analysts would mistakenly act as if something more than just the use of words grounds their disputes, and this would be the source of their method's mere verbalness.

This may not be a correct way to view the Canberra plan however. I doubt that many Canberra planners take themselves to be uncovering anything more than our conceptual structure with their analyses (and of course, if those analyses are realised). Then the questions they would be seeking to answer would be better described as *wholly* rather than *merely* verbal. As seen in the previous chapter, this is not necessarily a vice. Lexicographers and linguists would engage in these

disputes and questions as part of their role, and we would not say they are engaging in something nonsubstantive. Perhaps then the conceptual analyst should be thought of as a glorified lexicographer.

However, the objection may still stand that there is perhaps not much substantive that hangs on the question of what the folk concept is. Chalmers states:

The language and psychology of philosophy are important topics in their own right. But there is much more to philosophy than the language and psychology of philosophy, and great care is required in moving from the latter to the former . . . To see the point, note that the mere fact that existing words like 'know' or 'intentional' or 'see' behave in a certain way does not suffice to settle substantive disputes about epistemology, action, or perception. After all, views based on these data may differ only verbally from views on which 'know' or 'intentional' or 'see' pick out something else and that endorse apparently different first-order claims that use these words. (2011: 541).

The Canberra planner may or may not be happy with their work not being particularly non-linguistically substantive in this way. But regardless, when endorsing a particular near-realiser of a folk-concept, when others are in the vicinity, the Canberra planner would perhaps still face the charge of arbitrariness. They would be able to appeal to closeness to the folk theory in favour of one analysis if there are others which are equally close.

I think Carnapian explication gives the conceptual analyst a way out of this charge. With Carnapian explication the choice between different near-realises (or even far-realises) would be less arbitrary than if our primary purpose, in analysing concepts, was descriptive. This is because with the Carnapian view of analysis, we have certain criteria which different analyses will fulfil to differing extents. Even if two explicata fulfil the similarity criterion to the same extent, they differ when it comes to fulfilling the other criteria. What under the Canberra system would be an arbitrary choice between near-realises, could become a more substantive one when we judge their simplicity, exactness, fruitfulness, and similarity to the explicandum. The fact that under Carnap's system we are not merely limited to near realisers—as we can balance similarity

to the explicandum against the other criteria—also gives us a greater amount of options when it comes to finding an explicatum. We may not find a near-realiser of a concept X that can perform the role we want it to, but a far-realiser may do. This is opposed to the Canberra plan, where the lack of a near-realiser entails eliminativism about the folk-concept. In this way, for the Carnapian, the question of “what is X?” is not merely verbal, as for them the question can be paraphrased as “Which explicatum fulfils the four criteria to the greatest extent?”

## Problems for my account

My account is perhaps not without its problems and I anticipate several objections. I divide them into objections to Carnapian explication itself and objections to my use of it.

### Objections to explication

#### *Strawson's Objection*

I emphasised early on in this chapter the importance of defining the subject. When we ask questions about beliefs, about knowledge, about free will, etc. our questions are presumably framed within natural language and are concerned with these things according to the ordinary conception. If we show that, for example, free will exists according to a rather different conception – but one that meets Carnap's four criteria – then can we really say that we have solved the original philosophical problem? P. F. Strawson's criticism centres around the claim that

explicating a concept of philosophical interest by replacing it with a more scientific concept does not help us solve any philosophical problems. He states:

It seems *prima facie* evident that to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant—is a sheer misunderstanding, like offering a text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (1963: 505).

It seems to require no argument to show that, in most cases, either the operation (scientific language replacing non-scientific language) would not be practically feasible or the result of attempting it would be so radically different from the original that it could no longer be said to be fulfilling the same purpose, doing the same thing. (*ibid*)

If these things are true, it follows that typical philosophical problems about the concepts used in non-scientific discourse cannot be solved by laying down the rules of use of exact and fruitful concepts in science. (1963: 506).

This is not merely Jackson's worry about changing the subject. This is a worry about changing the subject in a particular way, i.e. replacing a folk concept with a scientific one. The purposes of a scientific concept are presumably different from that of a folk one; scientific concepts are highly specialised whereas folk concepts can be used for far wider purposes. For example, the language of morality may not have much use in science it is certainly very useful in everyday discourse. This puts doubt upon Carnap's response:

A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But for certain specific purposes, special tools are more efficient, e.g., chisels, cutting machines, and finally the microtome . . . [Strawson's] thesis is like saying that by using a special tool we evade the problem of the correct use of the cruder tool. But would anyone criticize the bacteriologist for using a microtome, and assert that he is evading the problem of correctly using the pocketknife? (Carnap, 1963: 938-939).

Carnap's response to Strawson is that natural language is broadly useful, but for some purposes we need to use a more specialised system. I am assuming (in my

terminology) he means that for some purposes our folk concepts (the explicanda) are inadequate, and that we would need to replace them with more specialised concepts (explicata).

However, to expand upon Carnap's analogy, although both the pocketknife and the microtome are cutting tools, in our investigation we may not necessarily be interested in cutting. Strawson does not merely want to cut using the pocketknife, he wants to investigate how the knife is used, the history behind the knife, the connections it has to our other tools, etc. Strawson may want to investigate our concepts and describe them because they are interesting in themselves. This coheres with his focus in *Individuals* on *descriptive* rather than *revisionary* metaphysics. He states: "Descriptive metaphysics is content to describe the actual structure of our thought about the world, revisionary metaphysics is concerned to produce a better structure" (1959: 9). Descriptive metaphysics, according to Strawson, is different to conceptual analysis in that the aim of descriptive metaphysics is far more general, "aiming lay bare the most general features of our conceptual structure" (1959: 9-10); this metaphysician does not aim at making explicit the application conditions of one concept, rather exposing the whole structure underlying it. Regardless the aim is the same: capturing the structure of our thought in some way.

I see no problem with this other than the arguments against traditional conceptual analysis already raised. The way we use the crude pocketknife, or its connections to other tools, may be interesting in themselves. However I am interested in finding tools which are better for cutting (cutting up reality that is!); it would be strange to say that this is not a worthy aim as well. I want to find fruitful and exact concepts that can play certain roles; the investigation into our language on its own is interesting, however it may not give us an answer to location problems (after all, it is not beyond the realms of possibility that our concepts may be false, or inadequate in some way) which is what I am interested in finding. The focus that Strawson puts on scientific language may also be unfair. As Carnap points out, all that is *essential* for the explicatum is that it be more exact than explicandum. In this way we would not replace a pocket knife with complex machinery, but perhaps with a different pocket knife with a shaper blade, a firmer handle and instructions for use. Many concepts of philosophical

interest would perhaps not be replaced by scientific ones. However, this does not mean that we could not replace them with more simple, more exact, more fruitful concepts.

## Objections to my account

### *Can explication really be considered a form of analysis?*

One objection that might be raised against my project is that explication is not a form of analysis. Consider analytic truths. Analytic truths are commonly taken to be true in virtue of meaning. For example, “All joeys are infant marsupials” would be an analytic truth on this picture; it follows from the definition of the target concept. This model of analytic truth suggests the decompositional picture: An analysis of a concept would only give us what was already part of that concept in the first place; what would follow from the meaning of the concept (this could of course include the definition itself, which seems to be what we usually aim at articulating when we perform conceptual analysis).

Explication does not seem to fit this conception of analysis. Let’s say that we explicated FISH into PISCES. Now as the only criteria for falling under FISH is being a water-dwelling animal, an explication of this concept that excludes otters, crabs, squids, etc. does not follow from the original concept, rather we explicate in this way because of other considerations (e.g. fruitfulness). C. H. Langford’s statement that an analysis “states an appropriate relation of equivalence” between what is analysed and that which does the analysing also expresses a similar thought. He goes on to state that if they “do not have the same meaning, the analysis is incorrect” (1942: 323). However, I would say that an acceptable explicatum would not have to have the same meaning as its

explicandum. So how then can we consider explication a form of analysis?

Beaney is one commentator who talks about explication as a form of analysis. He states, for example that “In his later work Carnap talks of analysis as ‘explication’” (2012). I think we have a better reason to think of explication as analysis than simply a deferral to certain peoples usage of the word ‘analysis’ however. My reason for thinking that explication is a form of analysis is because it can play a very similar role to what we want analysis for in philosophy, as we have shown. This role is essentially the making possible of serious metaphysics; the solving of location problems. When we factor in the fact that analysis has been conceived of in quite a few different ways, not merely as decomposition, there seems to be little reason for explication not to be counted as a form of analysis. Perhaps we can explicate analysis so that it includes explication in this way. Explication should be counted as a form of analysis because it does the same work we want analysis to do (it is similar to the explicandum), and it gives us a way to solve location problems even if the target concept has open texture (it is fruitful). This explication of analysis does not rely on circular reasoning. My conclusion that explication should be counted as a form of analysis is not contained in the premises. In order to perform an explication of analysis we do not need explication to be a form of analysis itself. So arguing via explication that analysis should include explication is not circular.

This issue may simply boil down to a mere verbal dispute over what 'analysis' means. The point of this essay could simply be that there is a method rather like conceptual analysis which we can use to do a lot of the work we want conceptual analysis to do and which avoids the arguments against conceptual analysis. We should not be worried about whether we can call explication a form of analysis. All that matters is whether it can help us in serious metaphysics in a similar way to conceptual analysis and whether it can avoid the arguments directed against conceptual analysis. I believe that I have shown that explication can do both of these things.

## Conclusion

In summary, my argument is, at its most basic, this: Williamson's argument aim to show that we should not expect a successful analysis because they always seem to be open to counterexample. Essentially his argument depends on putative analyses being defeated by type-1 counterexamples. A type of the conceptual analysis that's primary purpose is to analyse the folk conception of a concept is open to such counterexamples. These counterexamples are often fatal to such an account. Against this type of conceptual analysis Williamson's argument may work. However, if we change the purpose of our analysis then we avoid these sorts of counterexamples. Carnapian explication allows us to do this, and so without such powerful defeaters to an account we could perhaps achieve what could be considered successful analyses; in this case such analyses would be explicatums that satisfy Carnap's four criteria. As we can produce successful analyses on this picture, Williamson's argument does not seem to hold up.

Carnapian explication can also get us over the deflationary objection. It is true, that conceptual analysis, when practised in the right way (i.e. without getting ahead of ourselves) is not a merely verbal enterprise (though it may be wholly verbal). But, when analysing a folk-theory, we may run into the problem of equally near, but distinct realisers. Carnapian explication can help us get over this particular problem, as there are other criteria than similarity to the explicandum that the distinct explicatums may satisfy to different extents. For example, one may be more exact than the other. In this way, Carnapian explication makes the choice between analyses more principled, and less arbitrary.

## Bibliography

- Awodey, S. and Klein, C. (2004). *Carnap Brought Home: The View from Jena*. Illinois: Open Court.
- Bealer, G. (1998). "Intuition and the Autonomy of Philosophy" in DePaul & Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Rowman & Littlefield. 201-240.
- Beaney, M. (Eds.) (2007). *The Analytic Turn: Analysis in Early Analytic Philosophy and Phenomenology*. New York: Routledge.
- Beaney, M. "Analysis", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL=<<http://plato.stanford.edu/archives/win2012/entries/analysis/>>.
- Braddon-Mitchell, D., & Nola, R. (Eds.) (2009). *Conceptual Analysis and Philosophical Naturalism*. Cambridge, MA: MIT Press.
- Brandom, B. (2008) *Between Saying and Doing: Towards an Analytic Pragmatism*, Oxford: OUP.
- Cappelen, H. (2013) "Nonsense and illusions of thought" *Philosophical Perspectives*, 27, *Philosophy of language*, 22-50.
- Cappelen, H. and Hawthorne, J. (2009). *Relativism and Monadic Truth*, Oxford: OUP.
- Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.
- Carnap, R. (1963a). "P. F. Strawson on linguistic naturalism". In P. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 933–940). LaSalle: Open Court.
- Cassam, Q. (2009) "Can the concept of knowledge be analysed" in D. Pritchard

- and P. Greenough (eds.) *Williamson on Knowledge*. Oxford: OUP, pp. 12-30.
- Chalmers, D. (2011). "Verbal disputes" *Philosophical Review*, 120, 515–566.
- Chalmers, D (2012). *Constructing the world*. Oxford: OUP
- Chalmers, D. Manley, D. and Wasserman, R. (ed.) (2009) *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: OUP.
- Dennett, D. C. 1987. *The Intentional Stance*, Cambridge MA: MIT press.
- Dennett, D. C. 1991. 'Real Patterns' *Journal of Philosophy*, 87: 27-51.
- Gibbard, A. (2003), *Thinking How to Live*, Mass: Harvard.
- Hirsch, E. (1993). *Dividing Reality*, Oxford: OUP.
- Hirsch, E. (2005) "Physical-Object Ontology, Verbal Disputes and Common Sense" in *Philosophy and Phenomenological Research* 70, pp. 67-97.
- Hirsch, E. (2009). "Ontology and Alternative Languages." In *Metametaphysics: New Essays on the Foundations of Ontology*, ed. David J. Chalmers, David Manley, and Ryan Wasserman, 231–59. Oxford: OUP.
- Huvenes, T. (2012). "Varieties of disagreement and predicates of taste" in *Australasian Journal of Philosophy*, 90: 167–181.
- Huvenes, T. (2014). "Disagreement Without Error" in *Erkenntnis*, vol. 79, p. 143-154.
- Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: OUP.
- Jackson, F. (2005) "Ramsey sentences and avoiding the sui generis" in *Ramsey's legacy*, Oxford: OUP.
- Jackson, F. and Braddon-Mitchell, D. (2006) *Philosophy of Mind and Cognition: An Introduction, 2nd Edition*.
- Jenkins, C. S. I. (2014). "Merely Verbal Disputes" in *Erkenntnis*. 79 (1), 11-30.

- Justus, J. (2012) "Carnap on Concept Determination: Methodology for Philosophy of Science." *European Journal for Philosophy of Science* 2, 161-179.
- Kant, I. (2003) *Critique of Pure Reason*, Kemp Smith, Norman, (trans.) Hampshire: Palgrave Macmillan.
- Lewis, D. (1970) "How to define theoretical terms", *Journal of Philosophy*. 67, 427-446.
- Lewis, D. (1972) "Psychophysical and theoretical identifications". *Austrasian Journal of Philosophy*, 50, 249-258.
- Lycan, W. G. (1988) *Judgement and Justification*, Cambridge: CUP
- MacFalane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: OUP.
- Margolis, E. and Laurence, S. (ed.) (1999) *Concepts: Core readings*, Cambridge, MA: MIT Press.
- Moore, G.E., (1899) 'The Nature of Judgement' in Moore *Selected Writings*, (ed.) Thomas Baldwin, London: Routledge, 1993, 1-19.
- Nolan, D. (2005) *David Lewis*. Chesham: Acumen.
- Nolan, D. (2009) "Platitudes and metaphysics" in Braddon-Mitchell, D., & Nola, R. (Eds.). *Conceptual Analysis and Philosophical Naturalism*. Cambridge, MA: MIT Press, pp. 267-300.
- Platts, M. (1979) *Ways of Meaning*. London: Routledge and Kegan Paul.
- Price, H. (1988), *Facts and the Function of Truth*, Oxford: Blackwell.
- Price, H. (2003), "Truth as a Convenient Friction" *The Journal of Philosophy*, 100: 167-190. (Page numbers in text from Price (2011)).
- Price, H. (2011), *Naturalism Without Mirrors*, New York: OUP.
- Psillos, S. (1999) *How Science Tracks Truth*, New York: Routledge.
- Quine, W.V.O. (1951) "Two Dogmas of Empiricism" in *Philosophical Review*. 60,

20-43.

Quine, W.V.O. (1960) *Word and Object*, Cambridge, MA: MIT Press.

Rey, Georges, "The Analytic/Synthetic Distinction", *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2012/entries/analytic-synthetic/>.

Ridge, M. (2013), "Disagreement" *Philosophy and Phenomenological research*, Volume 86, Issue 1, pp. 41–63.

Smith, M. (1994). *The Moral Problem*. Oxford: Basil Blackwell.

Strawson, P. F. (1959). *Individuals*, London: Methuen & Co.

Strawson, P. F. (1963). "Carnap's views on constructed systems versus natural languages in analytic philosophy". In P. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 503–518). LaSalle: Open Court.

Strawson, P. F. (1992) *Analysis and Metaphysics: An Introduction to Philosophy*, Oxford: OUP.

Stevenson, C. L. (1944). *Ethics and Language*. New Haven: Yale University Press.

Stevenson, C. L. (1963). *Facts and Values*. New Haven: Yale University Press.

Waismann, F. "Verifiability". In Gilbert Ryle & Antony Flew (eds.), *Logic and Language (First Series): Essays*. B. Blackwell (1951).

Weatherson, B. (2003). "What Good are Counterexamples?". *Philosophical Studies* 115: 1–31.

Weatherson, B. (2009) "Conditionals and Indexical Relativism", *Synthese*, 166/2: 333–57.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, T (2009) "Replies to Kornblith, Jackson and Moore" *Analysis* 69 (1):125-135.