On Indexicals and Contexts

An Inquiry Into Modal Logic and Artificial Agents

Jonas De Vuyst

On Indexicals and Contexts

An Inquiry Into Modal Logic and Artificial Agents

Jonas De Vuyst

May 15, 2008

Abstract

In this paper we argue with Frege, against David Kaplan, that cognitive content, not the character of thought, triggers behavior. We also develop a logic of reconstructions to explain how agents can sometimes be cognitively blind to contradictions involving indexicals. Additionally, we argue that the character of indexicals varies with the genre of the reconstruction in which they are employed. This also implies that being mistaken about what one is reconstructing is a type of reference failure too.

We also research attributer and subject contextualism. We inquire how contextualism can help defeat skepticism and help solve other epistemological paradoxes. Finally, we investigate how epistemological contextualism can be incorporated into our logic of reconstructions.

Summary

Indexicals are terms such as 'I,' 'you,' and 'that vase over there,' whose meanings depend on context. David Kaplan—whose view is now orthodoxy—treats indexicals as descriptions whose referents are partly determined by the context in which the indexical is used. Thus the statement "That vase over there could have stood here instead" is not a contradiction because in the context in which this statement is made—presumably—there is a vase 'over there.' It is equivalent to saying that the vase that *actually* stands over there could have stood here instead. Another result by Kaplan is that although the statement "I am here now" cannot be uttered falsely, it is not necessarily true.

Kaplan calls the kind of meaning that is contextinvariant the character of a term or proposition. For instance, the character of the indexical 'I' might be 'the person using the word "I."' The kind of meaning that takes the context of use into account, he calls content. The content of the indexical 'I' would be a person who had used this word. Kaplan also applies this dichotomy to thought. According to Kaplan, the character of thought determines behavior: Thinking "I am on fire," will trigger different actions than thinking "He is on fire" because the characters of 'I' and 'he' differ. We think this is odd because if we were told "Do you know that guy?" our reply would depend on who we thought 'that guy' was. The semantic content of thought does not determine behavior either because what matters is not who 'that guy' is but who we think he is. Therefore it's the cognitive content of thought that triggers behavior.

Suppose a sophisticated setup of mirrors tricks you into thinking that you see two guys wearing a business suit and a diving suit, when really you're seeing just one guy wearing a bizarre costume that has the left half looking like a business suit and the right half looking like a diving suit. Suppose that in these circumstances you say "The first guy is wearing a business suit, but the other guy is not." This is clearly a contradiction as the 'first' and the 'other' guy are the same person. Kaplan's theory does not explain *how* one could come to believe such a contradiction though. We have developed a logic of reconstructions to fill this explanatory gap. Our theory models how people can, on a cognitive level, make mistakes about what an indexical refers to in some context. We

hold that reconstructions are anti-reflexive, acyclic, and not transitive. Additionally, we think they cannot fully model belief states.

Our logic of reconstructions has uses outside cognitive modeling too. In a historical text a writer might write "Now Siddharta leaves the palace." By making the character of indexicals vary with reconstruction genre, we can explain why the term 'now' in that text does not refer to the time at which it which it was authored—we reckon that historical texts define the character of the indexical 'now' as the last time described. Here's another example of how indexicals are determined by reconstructions. Suppose Amanda is about to present a radio show. Her show is announced as a news show, yet the first thing Amanda says is "Now the French are invading England!" as Amanda thinks she is supposed to present an episode of her radio program on twentieth century history. Evidently it is false that anyone is *now* invading England. We believe Amanda thought the sentence was true because she was mistaken about what she thought she was reconstructing.

We have also collected evidence on how to incorporate the two major families of epistemological contextualism—viz. attributer and subject contextualism—into our logic of reconstructions. The difference between attributer and subject contextualism is that the former usually states that the truth value of "S knows that p" varies with the conversation in which the knowledge ascription is evaluated whereas the latter holds it varies with the context of S.

Contextualism offers several devices to defeat skepticism. For instance, one important skeptical thesis states that we cannot know that we saw a zebra at the zoo because we could have been fooled by a painted mule instead. Fred Dretske replies that a subject really can know that he saw a zebra if he manages to rule out all relevant alternatives. Thus, usually a subject knows he saw a zebra if he saw an animal that looked like a striped horse. On the other hand, if the subject knew the zoo had tried to pass off a mule for a zebra in the past, he must make sure that the zebra is not a mule to know it's zebra. Other defenses against skepticism include the position that knowledge of propositions may rest on assumptions as long as one is willing to defend those assumptions when they are (legitimately) criticized (Antonia Barke and Michael Williams). Particularly Williams's contextualism impresses us because he has a mature account of contexts that takes semantic, conversational, epistemic, economic, and objective factors into account.

The lottery paradox makes for another interesting challenge. This paradox has it that lottery players will generally claim to not know that their tickets will lose, yet they'll also say they know they'll never be multi-millionaires. Stewart Cohen thinks the explanation for this behavior is that people will generally only notice the small chance of winning once standards for knowledge are raised. A related paradox is that when we're waiting for a

train and someone asks us if that train will arrive in ten minutes, we will readily answer "Yes" if the train normally does arrive at that time. Should the other guy, however, mention that it's very important that he's not late then we might have to concede that we actually don't know what time it arrives because we didn't take possible delays into account yet.

Contents

1. Introduction

1.1. PROBLEM STATEMENT

This paper consists of two parts. In the first part we will investigate the nature of *indexicals*—that is, phrases such as 'I,' 'she,' 'that chair over there,' which depend on context for their contents[1]. We will start off by drawing on David Kaplan's theory of indexicals to set us up with a well known conceptual apparatus. We will then explore the strengths and weaknesses of Kaplan's theory. Special attention will go to semantic and cognitive aspects to thought. Particularly, we will focus on the curious phenomenon that Kaplan's theory seems to predict that via indexicals agents can come to believe for certain objects x that $P(x) \wedge \neg P(x)$. Kaplan does not give an explanation for *how* one can believe such a contradiction so we will set out to create *a logic of reconstructions* in which such contradictions do not unavoidably lead to inconsistency on a cognitive level. Additionally we will investigate some formal properties of our logic. Also, we will inquire about the philosophical link between belief states and reconstructions. Finally we will use reconstructions to account for some uses of indexicals in English that go against some of the predictions made by Kaplan.

In the second part of this paper we will first research contemporary positions in episte-mological contextualism. Early on we will find that there are two important contextualist families in epistemology—namely *attributer* and *subject contextualism.* Some of the most important proponents of attributer contextualism are Stewart Cohen, David Lewis, and Keith DeRose. Subject contextualism is defended by Michael Williams and Antonia Barke (among others). We will investigate how contextualism deals with skepticism and how it attempts to explain for some other paradoxes. Extra attention will go to Williams's epistemological theory because of its apparent complexity. We will end this report by outlining how our logic of reconstructions might be combined with either attributer or

---

[1]Throughout this paper we will use 'contents' as a plural for 'content.' We will also use the term '*a* content' when writing about the—presumably standalone—existence of a proposition's content.

subject contextualism.

We have selected the topics treated in this paper with an eye on *artificial intelligence* (AI) and *artificial agents*. For instance, we believe that indexicals are important for efficient communication—be it in natural language or between artificial agents. Also, we will dedicate a large part of this paper to cognitive aspects of indexicals. Finally, we believe epistemological contextualism might be of interest to AI researchers—especially those concerned with the problem of planning—because of its concern with defeating skepticism and solving other epistemic problems *on a practical level*.

1.2. MOST IMPORTANT SOURCES

We will use the article 'Indexicals' from the *Stanford Encyclopedia of Philosophy* (Braun, 2007) to get us started on indexicals. Kaplan will be our main guide on this topic, though, as he's behind the foundations of indexicals used in contemporary logic and analytical philosophy (1989*a*, 1989*b*).

Concerning epistemology we will use two main sources. The first is Williams's book on epistemology, which we will use for information about general epistemology as well as on Williams's own contextualist theory (2001). The second source is *Erkenntnis*, volume 61. This issue of the famous journal was dedicated to epistemological contextualism and as such provides us with many useful articles.

Hughes and Cresswell provide us with our primary resource on modal logic (1996). Additionally we will make use of the article 'Modal Logic' in the *Stanford Encyclopedia of Philosophy* (Garson, 2008).

2. Indexicals

Indexicals are expressions—usually single words—whose *contents* and *referents*[1] depend in part on the *context* of the speaker. Some paradigmatic examples are 'I,' 'here,' 'now,' 'today,' 'he,' 'she,' and 'that.' In a tradition that traces back to Kaplan, the meaning of an indexical abstracted from its context, is often termed its *character* or *linguistic meaning*. As such the word 'I' has a single character but different referents when first you and then we use it. (Braun, 2007, up to § 1.1)

A second—coarser grained—distinction due to Kaplan is that between *pure indexicals* and *true demonstratives*. 'I' is a pure indexical—it always refers to the speaker. Similarly, 'today,' 'tomorrow,' 'actual,' 'present,' 'here,' and 'now' seem to be pure indexicals as their contents too appear to be derivable directly from the speaker's context. Contrast this to the true demonstratives 'he,' 'she,' 'his,' 'her,' and 'that,' which seemingly demand of the speaker that he indicates which person or object he is referring too. (Braun, 2007, § 1.3)

It has been suggested that many more words should be subsumed under categories of indexicals. Now, obviously, there also exist variations on Kaplan's indexicals: 'we,' 'ours,' 'they,' 'theirs,' 'these,' and 'those.' But, more pertinently, various philosophers have suggested we should regard as indexicals phenomena as diverse as these:

(1) Words and morphemes that indicate tense. From context to context, these may refer to different times or time intervals.

(2) 'Come,' 'go,' 'left,' and 'right' look like they similarly depend on context;

(3) Modal expressions. The *kind* of modality employed when we use the words 'possibly' or 'necessarily' seems to vary with context. At one time and place we may be speaking in terms of metaphysical possibility when we say "It is possible that ...,"

---

[1] The terms 'content' and 'referent' are not synonymous. Braun (2007) uses the term 'content' for both propositions and indexicals. Use of the term referent, though, is limitted to indexicals and other singular terms. However, aside from this grammatical difference one can think of these terms as if their meanings were interchangeable.

while on another occasion we might use these same words when discussing epistemic possibilities;

(4) Counterfactuals. These are phrases such as "If p were the case, then ...," when p is *in fact* not the case. According to Lewis the range of possibilities that we need to consider when evaluating such statements, is context-dependent;

(5) Propositional attitude verbs, such as 'believe' and 'know.' Mark Richard thinks that when we say that "S believes p," we really mean to express that S would accept q, where q is a contextual translation of p. Thus if Lois Lane does not suspect that Clark Kent is Superman, the truth value of "Lois believes that Clark Kent can fly" may shift from context to context because p can mean different things. On the topic of knowledge, *conversational contextualists* (CC) Cohen, DeRose, and Lewis think that although knowledge claims such as "S knows he has a hand" are false when certain skeptical arguments are salient, they may nevertheless be true before such arguments are considered or after they are forgotten. For these philosophers such knowledge ascriptions are always to be seen in the light of *standards set by the conversation*; (See also Brendel and Jäger, 2004)

(6) Comparatives such as 'rich,' 'tall,' 'large,' 'heavy,' 'hot,' and 'fast;'

(7) The word 'local.' Contrast the expression "We went to see Tarantino's movie at a local theater" to "Everyone concerned with the environment watched Al Gore's movie at a local theater." The first sentence commits us only to the existence of exactly one theater, while the second sentence implies that Gore's movie may have played at many theaters, but might as well have played at no theaters at all; and

(8) Vague terms. What makes a heap? How many hairs must a person have to escape the predicate 'bald?' Philosophers will habitually test replies to these question for the *sorites paradox*. This paradox relies on the argument that if n rocks make a heap, then surely $n-1$ rocks must make a heap too (for what difference does removing one rock make?). But if $n-1$ rocks make a heap then by the same argument $n-2$ rocks must be a heap too. Regardless of the size of the initial heap, tirelessly repeating this argument will yield the absurd conclusion that a single rock too constitutes a heap. Some theorists believe this problem can be (dis)solved by allowing any vague expression to be true in some contexts, while simultaneously being false in others. (See, e.g., Gert, 2007)

(For these more contentious indexicals, cf. Braun, 2007, § 1.4)

In this chapter we will focus on the sort of indexicals identified by Kaplan. We shall explain what philosophical and logical problems they bring about, and will propose a remedy for these problems.

2.1. PURE INDEXICALS AND TRUE DEMONSTRATIVES

Suppose that while you're reading a philosophical paper you suddenly hear a scary voice. It says:

> I see you.

You take a look around the room but no one is there with you. Just as you're about to shrug off what you think you heard, the voice repeats itself:

> I see you.

You clear your throat, and ask: "Who, me?" The room is silent for the briefest of moments. You speak up again: "Who said that?"

We break off this mystery story here for we're afraid we're not much of fiction writers. Our little digression wasn't without purpose though. Recall that for Kaplan there (basically) are two sorts of indexicals. We can distinguish them readily in the message "I see you." The pure indexical here is the word 'I;' the true demonstrative is the word 'you.' Notice also that our story had a bit of a confusing side to it—two unusual questions were raised: (1) Who is 'I' and (2) who is 'you?' As language users, filling in these blanks normally comes natural to us. But here this skill of ours broke down. Below we shall exploit features of this breakdown to learn about indexicals.

2.1.1. Deducing Content

"Who is 'I?'" This question sounds all wrong. It sounds so wrong that perhaps for a moment you thought a grammatical error crept in there. But allow us to earnestly answer the question: When you say 'I,' you're referring to yourself. Similarly, when *we* use the word 'I,' we mean us. Thus, the function of 'I' to its referent would appear to be interpretable by the operation that maps 'I' to its user.

Let us put this property in formal terms. Suppose that we have at our disposal a number of agents. These agents have belief systems whose implementation are based on formal predicate logic. We say that for every agent a with belief system $\mathcal{A}$,

$$a \text{ believes that } \alpha \iff \mathcal{A} \models \alpha$$

Now, a formalization of the hypothesis above entails that, for every agent a with belief system $\mathcal{A}$, if $\mathcal{A}$ has a in its domain, available as e (be it a constant or a function in $\mathcal{A}$), and if 'I' is transliterated in $\mathcal{A}$ as i,

$$\mathcal{A} \models \alpha \iff \mathcal{A} \models [e/i]\alpha$$

where $[e/i]\phi$ is the result yielded by an indiscriminate substitution of e for i in α. Analogous schemata can be constructed for other sentence usages by introducing logs of what a has asserted, imagined, and so on. For instance, if for every agent a with belief system $\mathcal{A}$, Santa Claus keeps a log $\mathcal{L}$ of lies told by a, then

$$\alpha \in \mathcal{L} \implies \mathcal{A} \not\models \alpha$$

and thus, if a knows its name is e, and uses symbol i for 'I,'

$$\psi \in \mathcal{L} \implies \mathcal{A} \not\models [e/i]\psi$$

The totality of these schemata—the remainder of which we will omit here—formalizes the suspicion that the name 'I' always refers to its user.

The formulas above suggest that it is not an empirical matter that 'I' refers to its user; rather, if our notation is correct then this is always the case by way of analytical truth. As to explaining why this might be so, Kaplan's notion of contexts may prove helpful. Kaplan thinks an agent's context consists at the least of an ordered quadruple,

$$\mathcal{C} = \langle a, l, t, w \rangle$$

where a is the agent, l is its current location, t is the current time, and w is the context's world. These objects are best familiar to us as the contents of 'I,' 'here,' 'now,' and 'actual.' The term 'actual' should here be understood as an antonym to 'counterfactual'—i.e. it designates that which is the case, as opposed to what could have been the case if such and so were different. (Braun, 2007, § 3.2)

Kaplan has developed a modal logic to go with his notion of contexts. Unlike our own formal description of the word 'I,' however, his logic does not model belief systems and whatnot. Thus in Kaplan's logic there is no privileged user of the word 'I' and the actual and possible worlds that it describes are, in a way, disembodied worlds. The approach of abstracting away the epistemic perspectives of agents is not without its advantages though. Most importantly, it affords a logical picture of some of the problems of communicating

sentences that contain indexicals. For consider that it might very well be possible to resolve indexicals into their referents before the statements that they are embedded in are stored in an agent's belief system. If this is true for all indexicals then from an engineering point of view it might possibly also be desirable to offload the complexity of dealing with indexicals from the belief system algorithms onto a separate subsystem. Therefore, even though the goal of our inquiry into indexicals is to find out how they affect the design of artificial agents, we will presently follow Kaplan in his use of logic from a god's eye point of view.

Recall that we suggested that it is an analytical fact that 'I' refers to the user of this word. Now, does this imply that the sentence

$$\text{I am here now} \tag{2.1}$$

cannot possibly be false? Kaplan believes that in one way it cannot be false, while in another it can (1989a, § VII, XI, XIII). In an important sense this sentence cannot be false because it simply cannot be *stated* falsely. However, this does not mean that we could not have possibly been anywhere but here. So it's false that

$$\square \text{ I am here now.} \tag{2.2}$$

For instance, we would already have been out eating right now if our friends weren't stuck in traffic. For Kaplan the confusing intuitions that in one way (2.1) cannot be false, while in another it can, stem from a confusion between context of use and *circumstances of evaluation*. On Kaplan's view, each sentence has a character as its distinguishing feature. When a sentence is placed in a context, as happens when it is uttered, the sentence's character and the context determine a content. Evaluating a content in the actual or in some counterfactual circumstance yields a *truth value*. Where indexicals come in is that their referent, like the overall content of a sentence, is determined by character and context. Thus, in (2.1) the word 'I' always refers to us—regardless whether you or us consider the statement. Likewise, in that same statement the term 'now' refers to the moment we (actually) wrote the phrase down, regardless of the circumstance in which it is evaluated. Hence what gives (2.1) the appearance of unavoidably being true is that it cannot be false *in the circumstance of its context* (and that's also true when someone else speaks the same words). Conversely (2.2) is false because in this sentence we also demand that (2.1) upholds in all possible counterfactual circumstances (and in some of those we are not here). Hence *the rule of necessitation* ($\vDash \alpha \implies \vDash \square\alpha$) should be rejected.

2.1.2. Dthat Girl In the Blue Sweater

Pretend we are having a conversation in real life. A girl wearing a blue sweater walks by. Seemingly for no good reason we comment that the girl in the blue sweater could have worn a green sweater instead. We trust that you agree with us on this one. On the other hand, it does seem tautologous to say that *a* girl in a blue sweater wears a blue sweater. What sets these sentences apart? Going by Kaplan, the formal answer is that the first sentence implicitly makes use of the *dthat operator* while the second does not (1989*a*, § XII, XIV, XV). Dthat[α] is a device for making *demonstrations* that *directly refer* to some content. Tentatively it could be used to define 'I' as follows:

$$\text{'I'} =_{\text{def}} \text{dthat[the person using the term 'I'].} \tag{2.3}$$

The girl's freedom to not wear a blue sweater can be expressed unambiguously as follows:

$$\Diamond \neg (\text{dthat[the girl wearing the blue sweater] wears a blue sweater).} \tag{2.4}$$

In (2.3) and (2.4) the descriptions 'the person using the term "I"' and 'the girl wearing the blue sweater' are the character of dthat demonstrations. The content of (2.4) might turn out to be equal to

$$\Diamond \neg \text{Laura wears a blue sweater.}$$

At least it might be supposing that Laura walks by and is wearing a blue sweater.

Notice how the indexical 'the girl wearing the blue sweater' is here resolved into the name 'Laura' on the basis of what is actually the case here and now. This happens despite the dthat formula being embedded in an expression that is within the scope of the $\Diamond$ operator. One way to grasp this is to picture expressions such as (2.4) being evaluated in two passes: (1) On the first pass dthat formulas are resolved into direct refererences; (2) on the second pass the truth value for the expression is computed. Of course, as discussed above, pass (1) takes place in the context wherein the sentence is uttered, whereas in (2) we may also compute the truth value for counterfactual circumstances.

2.2. COGNITIVE SIGNIFICANCE AND OBJECTS OF THOUGHT

For the purpose of explaining Kaplan's account of indexicals, we took a step back from discussing belief systems. We now return to discussing some of the aspects of indexicals that have implications on agent design.

2.2.1. Kaplan's Semantics of Thought

Kaplan brings up the following sentence (1989a, § XVII):

$$I \text{ am getting bored.} \tag{2.5}$$

An interesting question comes with this statement: When you and us say these same words to ourselves, do we think the same thing? It turns out that in one important way we don't—namely in the manner that you might be getting bored just when we're having the times of our lives. What happens here is that the *objects of thought* differ when you and us think proposition (2.5). Now, of course the reason these objects of thought are different is that the indexical 'I' has different referents between both thoughts. Kaplan believes this suggests that objects of thought are simply contents. That's an interesting way to look at it because it allows Kaplan to postulate a new aspect to thought, which he calls the *cognitive significance of an object of thought*, and which he identifies as the character of what we say to ourselves. When two people think (2.5), it is the cognitive significance of their thoughts that is identical. It is in terms of this cognitive significance of an object of thought that we speak of psychological states. It is also the cognitive significance of a thought that explains our behavior. Kaplan offers some examples to illustrate this latter point. For instance, suppose we're mistaken about the date and believe today is January 4 when really it's January 6. In this counterfactual situation we might hope to make some deadline or other by January 5, whereas if someone asked us to do something by 'yesterday' (i.e. also January 5) we would not think for a second that this was possible. Here's a second example: Suppose we're walking through a crowded street when, in the window of a shop, we see the reflection of a man without pants.[2] Depending on whether we think

$$He's \text{ not wearing any pants} \tag{2.6}$$

or

$$I'm \text{ not wearing pants,} \tag{2.7}$$

our reaction might be either to laugh or to pinch ourselves to see if we're really awake.

Kaplan's distinction between the objects of thought and their cognitive significance

---

[2]In Kaplan's original example the man's pants are on fire but we've never had *that* dream.

makes a good starting point for further discussion. Most importantly it provides us with vocabulary for describing the borderline between semantics and cognition. Kaplan's account—as he repeatedly stresses—deals only with the semantics though. This may sound odd because one might sooner expect a theory of thought to deal with cognition than with semantics. So let us first investigate what this means exactly. We already discussed Kaplan's semantics of indexicals as they apply to the objective world. We believe that what Kaplan aims to achieve here is to reuse his theory about indexicals to explain the *semantics of thought*. That is, he wants to explain what it means in objective terms for a subject to have a certain thought. Let me clarify this by returning to the example about the pants. Consider the additional sentence:

$$\text{Jonas is not wearing any pants.} \tag{2.8}$$

One might think that the *thoughts* that go along with (2.7) or (2.8) are identical. After all, in the context in which we entertain these thoughts, there is no room for doubt about what the indexical 'I' and the name 'Jonas' refer too. Nevertheless, as we just saw, according to Kaplan two thoughts having the same objects may yet trigger different behaviors if their cognitive significance differs. Therefore, Kaplan appears to contend, the psychological meaning of thoughts must be taken to vary with their cognitive significance.

What's odd about Kaplan's theory, though, is that it ignores the problem of perception in its entirety. Suppose we think we see someone who owes us money. If we're friendly with this person we might yell something at him like

$$\text{Hey, you there, you owe me twenty euros!} \tag{2.9}$$

Now if it turned out we were mistaken about who we thought we saw then this statement would obviously be false (not to mention rude). In fact, its falsehood would be so obvious that our error could not credibly be explained as something being wrong with our reasoning at the time. But how can we account for it then? One way out might be to say that we mistakenly believed a fact such as

$$\text{dthat[you there]} = \text{Daniel.}$$

This seems to explain *what* went wrong. It still is not clear, however, *how* someone could mistakenly come to believe something like this on Kaplan's semantic theory. Supposedly the *how* part of our inquiry falls outside the scope of semantics, but considering (2.9) again we are still left to wonder how a thought could mistakenly turn out to claim that one is

owed money by a complete stranger. Rather, it intuitively seems that, in thought, one ascribes properties only to concrete individuals and not to their names or to descriptions of them. Only by deduction do we appear to arrive at sentences such as (2.9). From this cognitive point of view Frege seems have been on to something when he wrote:

> Now everyone is presented to himself in a special and primitive way, in which he is presented to no one else. So, when Dr Lauben has the thought that he was wounded, he will probably be basing it on this primitive way in which he is presented to himself. And only Dr Lauben himself can grasp thoughts specified in this way. But now he may want to communicate with others. He cannot communicate a thought he alone can grasp. Therefore, if he now says 'I was wounded', he must use 'I' in a sense which can be grasped by others, perhaps in the sense of 'he who is speaking to you at this moment'; by doing this he makes the conditions accompanying his utterance serve towards the expression of a thought. (Frege, 1997b, p. 333)

Kaplan, from whom we borrowed the quotation above, offers a reductionist interpretation to the semantics of this text (1989a, § XVII). He believes it is simply the cognitive significance (the character or presentation) of 'I' that makes for the particular and primitive way in which Dr. Lauben is presented to himself. We reckon this is partially correct: The fact that

$$\text{dthat[the object I'm considering]} = moi \tag{2.10}$$

does tell us that the object we're thinking about is special to us in a manner that only affects ourselves. However, when we try to think of ourselves via the indexical

$$\text{dthat[the person sitting in this chair]}$$

We are no less disposed to act on (say) a feeling of hunger than we are disposed to satisfy the appetite of 'I.' Once we realize that it's us that's sitting in this chair, the manner in which we are presented to myself matters no more (if indexicals indeed make for a difference thought at all). Mind you, this point is primarily about cognition, not semantics. Nevertheless, if we are correct that once someone realizes he is thinking about himself, it is impossible to think of himself in abstract terms such as 'the guy's whose reflection I am looking at' or 'the guy sitting in this chair' then Kaplan's theory will have to be modified to explain for this behavior. That's if we want to keep his theory anyway—perhaps we can come up with a better account of the semantics of thought.

David Braun cites an interesting challenge for theories about the semantics of thought (2007, § 4.1–4.2). Suppose Jeff is wearing a bizarre costume that looks like an ordinary business suit on the left side and like a diving suit on the right side. Imagine further that you meet Jeff. Due to the way he's standing, however, you only directly see his left side. To the right of Jeff is a mirror, though, and in it you see his right side—but not his face, so you're unaware the guy in the mirror is Jeff. In this unusual situation you remark:

$$\text{You (addressing Jeff directly) are wearing a business suit,} \qquad (2.11)$$

$$\text{but he (pointing at the mirror) is not.} \qquad (2.12)$$

Braun believes that on Kaplan's view we should take you to believe a contradictory proposition in this situation.[3] He adds that because it's implausible that anyone would believe such a blatant contradiction, many philosophers believe Kaplan's theory is incorrect. Let's look at this more closely though. Recall that for Kaplan there are two aspects to a thought (1989a, § XVII). It is the cognitive significance of the objects of our thoughts that explains our actions, while the objects of thought is what our thoughts say about the world. If we may interpret making contradictory statements as a kind of action then the explanation for this action may be the innocuous mistake of not realizing that one has physically indicated the same person twice; it is to not realize that

$$\text{dthat['you' in (2.11)]} = \text{dthat['he' in (2.12)]}. \qquad (2.13)$$

Now in Kaplan's logic the truth value of sentences such as (2.13) becomes apparent once one evaluates the content of the sentence. In Braun's thought experiment (2.13) is true and that is why (2.11) contradicts (2.12). The contradiction, however, would be for *Jeff* to both be wearing and not wearing a business suit at the same time. There is, on the other hand, no *a priori* contradiction that 'someone indicated by you' would wear a given costume while 'someone indicated by you seconds later' was not wearing it. Thus, the argument repeated above does not falsify Kaplan's theory. Braun's example does, nevertheless, provide us with a second example of how Kaplan's semantic theory fails to dub as a theory of cognition. We are again left to wonder *how* an agent may fail to trace dthat-references. Particularly, Kaplan leaves unexplained what an agent may plausibly mistake the terms in (2.13) to reference.

---

[3]In this experiment it seems plausible that you would indeed believe both (2.11) and (2.12). Nevertheless, the propositions that Jeff is wearing a business suit and that he is wearing a diving suit are both false because his outfit is really more of a gimmick. Note, though, that (2.11) and (2.12) contradict eachother regardless of the costume he truly is wearing. It's the nature of this contradiction that intrigues Braun.

2.2.2. Difficulties In Modeling Reference Failure

It's an obvious fact that people can be mistaken about the referents of the indexicals they use in thought or in speech. Perhaps the most obvious mistake to make is to believe an indexical refers to object a, when really it references b (with $a \neq b$). This is what happens when you think you recognize someone, but don't. Another way to confuse a reference is to think an indexical refers to some object c, when c is actually but a figment of your imagination. This second kind of mistake can be modeled by making the indexical 'refer' (pretend-refer) to an object that exists in a possible world that represents the world as the—here overly imaginative—language user sees it. A third way to make an error about references is described in (2.11)–(2.13). What happens here is that an object is seen as two objects. Most possible world semantics cannot model this situation properly because they have the following properties:

$$x = y \supset \Box x = y, \qquad \text{(LI)}$$

$$x \neq y \supset \Box x \neq y. \qquad \text{(LNI)}$$

(Hughes and Cresswell, 1996, pp. 312–314) Still, with semantics that do not have these properties we might be able to model the third situation. Hughes and Cresswell describe two such systems, namely *counterpart theory* (due to Lewis) and *modal predicate logic* with *intentional objects* (1996, § 18–19).

Counterpart theory involves translating *well formed formulas* (wff) of modal predicate logic into an extensional *non-modal predicate logic.* Some of Lewis's metaphysical views led him to design counterpart theory such that objects only ever exist in a single world. They can, however, have zero, one, or more counterparts in other worlds. Counterparts of objects from world w_1 in world w_2 are specified by a *counterpart relation.* What's interesting about this approach is that a counterpart relation can be chosen that has some or none of the following properties: Reflexivity, transitivity, and symmetry. However, it cannot have *all* of these characteristics without losing its most intriguing feature—viz. that an object from w_1 that has no counterparts in w_1 aside from itself can have multiple counterparts in w_2. (Cresswell, 2004; Hughes and Cresswell, 1996, § 19)

Intentional objects make for a more straightforward approach. It is possible to extend an ordinary modal predicate logic to support intentional objects by modifying its valuation rules, the semantics of its assignment functions, and sometimes the structure of its models. Usually for a model $\langle W, R, D, V \rangle$—with W the set of worlds, R the accessibility relation, D the domain of objects, and V the valuation function—given assignment μ, every variable

x is assigned the value $\mu(x) \in D$. Atomic well formed formulas are usually evaluated like this:

$$V_\mu(\phi x_1 \ldots x_n, w) = 1 \iff \langle \mu(x_1), \ldots, \mu(x_n), w \rangle \in V(\phi). \qquad (V\phi)$$

To get intentional objects we need to modify the semantics of μ so that it assigns objects to variables in a world relative way. An easy way to do this is to give it a second argument: A variable x in world w is then assigned object $\mu(x, w) \in D$. The evaluation of atomic formulas needs to be modified accordingly:

$$V_\mu(\phi x_1 \ldots x_n, w) = 1 \iff \langle \mu(x_1, w), \ldots, \mu(x_n, w), w \rangle \in V(\phi). \qquad (V\phi')$$

In Hughes and Cresswell's modal predicate logic we'd also need a new definition for 'x-alternative,'

$$\mu \text{ and } \rho \text{ are x-alternatives} \iff \text{for every variable } y \neq x \text{ and } w \in W,$$
$$\rho(y, w) = \mu(x, w),$$

because they use this concept in $(V\forall)$:

$$V_\mu(\forall x\alpha, w) = 1 \iff V_\rho(\alpha, w) = 1 \text{ for every x-alternative } \rho \text{ of } \mu. \qquad (V\forall)$$

(Hughes and Cresswell, 1996, p. 243, § 18)

It is important that we mention that these modifications cause (LI) to no longer be valid. One other noteworthy change is that in this new system, (I2) will only be valid if α and β do not contain modal operators:

$$x = y \supset (\alpha \supset \beta). \qquad (I2)$$

Fortunately we can replace it by

$$x = y \supset (\phi x_1 \ldots x_n \equiv \phi y_1 \ldots y_n), \qquad (I2")$$

where for each $1 \leq i \leq n$, x_i and y_i are x and y respectively, or they are variables that are either both free or bound to the same object. A third, more interesting, result is that the semantics here outlined make the following schema valid:

$$\Box\exists x\alpha \supset \exists x\Box\alpha. \qquad (2.14)$$

14

(Hughes and Cresswell, 1996, § 17–18)

Traditionally (2.14) is considered highly undesirable. As Quine put it: "[I]n a game of a type admitting of no tie it is necessary that some one of the players will win, but there is no one player of whom it may be said to be necessary that he win" (1980*b*, p. 148). However, if we take there to be an intentional object 'the winner,' that has in each world for its extension the actual winner then it *can* be said after all that there is an individual that necessarily will win. The trick, then, is to allow quantification over any constructable intentional object—that is, over intentional objects with any cross-world extension whatsoever. (Hughes and Cresswell, 1996, § 18)

Alternatively, if we do not want (2.14) to hold—and we certainly don't depend on it to hold—we can extend the structure of our models yet further: $\langle W, R, D, I, V \rangle$. Newcomer I is the set of intentional objects that may be assigned to variables. Its elements $i \in I$ are functions from W to D. Thus intentional objects so understood yield an object when passed a world as a parameter. Accordingly, instead of assigning to x the value $\mu(x, w)$, we now get $\mu(x) \in I$ and assign $\mu(x)(w)$ to x. The evaluation procedure for atomic well formed formulas is modified similarly:

$$V_\mu(\phi x_1 \ldots x_n, w) = 1 \iff \langle \mu(x_1)(w), \ldots, \mu(x_n)(w), w \rangle \in V(\phi). \qquad (V\phi'')$$

Finally, we define x-alternatives as follows:

$$\rho \text{ is an x-alternative of } \mu \iff \text{ for every variable } y \neq x, \rho(y) = \mu(y).$$

As long as I does not contain all possible functions from W to D, (2.14) will not hold. Neither will (LNI), unless we reintroduce (LI) *and* make frames symmetrical. (Hughes and Cresswell, 1996, § 18)

Observe that whereas a counterpart relation may stipulate that objects b and c in w_2 are counterparts of a in w_1, extensions for intentional objects always contain exactly one object per world. Regardless, intentional objects *can* make for an alternative to counterparts. We will illustrate this point by way of example. Suppose this world was a better world and we had a bottle of beer standing in front of us. In this (counterfactual) situation we might have reflected that "If this world was a better world, we would have had two bottles of beer." In counterpart terminology we might say that the latter thought would describe two imaginary counterparts to the one bottle we supposedly had. Here's an informal description of this link between the three bottles of beer: If the world was

how we imagined[4] it to be, the beer we had would have been like the beers we thought up—e.g. it would have been cold, would have stood here and there, and it would have been two beers. Using the second variant on intentional object semantics described above, we need $i_1, i_2 \in I$ and $w_1, w_2 \in W$ to describe the same situation. If w_1 is the world in which we have one bottle of beer and w_2 is the world wherein we have two, then in w_2 let i_1 be the intentional object for the first bottle and let i_2 likewise represent the second beer. We then find that $i_1(w_2) \neq i_2(w_2)$ while $i_1(w_1) = i_2(w_1)$. (Cf. Hughes and Cresswell, 1996, pp. 31, 39–40, 357–358)

Note that counterpart relations cannot fully be described in terms of the equality and inequality of intentional objects. For whereas for $n > 1$ the counterpart relation that maps n objects from one world onto m objects in another world, with $n \geq m > 0$, can always be expressed like that, this technique breaks down when $m = 0$. So we need a different device to translate those counterpart relations that stipulate that certain objects from w_1 have no counterpart in w_2.

Because our logic uses a *fixed domain* for all worlds (up to this point), the concept that an object might exist in one world but not in another, has no immediate notation. There are basically two ways out of this—more basic—problem.

One solution is to interpret the domain of objects as the domain of all *possible* objects. A predicate E is then introduced that holds for x if and only if x really exists (obtains). The claim that there exists at least one human being might then be modified from

$$\exists x (H(x))$$

to

$$\exists x (H(x) \wedge E(x)).$$

(Garson, 2008, § 13)

Note that predicates do not say anything about intentional objects, but instead apply to objects from D.[5] Thus we cannot right away distinguish the case where an intentional object from I designates an object a from D and where a happens to not exist, from the

---

[4]Or imagined-imagined, as our starting point is a counterfactual situation.

[5]On the second approach to intentional objects, our stumbling block is that predicates always apply to objects in the domain D, not to intentional objects from I. On the first approach our obstacle is that predicates are applied to objects, not to variables. That is, for any assignment μ and world w, $\phi(x_1, \ldots, x_n)$ will be evaluated as $V_\mu(\phi x_1, \ldots, x_n, w)$, which by $(V\phi')$ is equivalent to $\langle \mu(x_1, w), \ldots, \mu(x_n, w), w \rangle \in V(P)$. As each $\mu(x_i, w)$ simply yields an object from D we again find that, ultimately, $\phi(x_1, \ldots, x_n)$ is evaluated in a context with no residue of the intentional object.

case where an intentional object from I does not designate anything—indeed, every such object must per our semantics always designate *something*. This isn't a problem for us though because an object having no counterpart—rather than a series of n intentional objects having m with $0 < m < n$ counterparts—can happen only in one of the following cases:

(1) There's an object $i \in I$ that designates $d \in D$ in the real world. The object exists but goes by unnoticed by Joe. We would still, however, preferably say that the object is a possible object in Joe's belief system. Therefore, formally, we should make e designate some object from D in the world representing Joe's beliefs. It could be that $d = e$ but it needn't be. The important thing is that there's an object that has predicates applying to it so that Joe can know of a possible red flower pot, which he has been told exists, but which he himself doesn't believe to exist because he hasn't seen it with his own eyes;

(2) Conversely, there could be an object $j \in I$ which designates $f \in D$ in the world representing Frank's beliefs. Frank is hallucinating though and so f doesn't exist in the real world. It would nevertheless still be best to make j designate f in the real world. That way the semantic world representing the real world could store the fact that the green ashtray that Frank thinks he sees, isn't real.

A second solution to the basic problem of codifying propositions that say that an object obtains in world w_1 but not in world w_2, is to use *world-relative domains*. (Garson, 2008, § 13) Ignoring intentional objects for the moment, we can get such systems by making models quintuples $\langle W, R, D, Q, V \rangle$, with Q a function from worlds onto domains $D_w \subseteq D$, and by also modifying $(V\forall)$ as follows:

$$V_\mu(\forall \alpha, w) = 1 \iff V_\rho(\alpha, w) = 1 \text{ for every x-alternative } \rho \text{ of } \mu \text{ such that } \rho(x) \in D_w$$
$$\text{and 0 otherwise.} \qquad (V\forall')$$

Well formed formula α is then valid if and only if $V_\rho(\alpha, w) = 1$ for all $w \in W$ and all assignments μ such that $\mu(x) \in D_w$ for every variable x. (Hughes and Cresswell, 1996, p. 275) In such setups we can usually test if t obtains by testing for $\exists x(x = t)$. (Garson, 2008, § 13)

Often world-relative domains are rejected because they have the side effect of invalidating the rule of necessitation. For instance, even though the theorem that if ϕ is true

for every x then ϕ is true for any handpicked object y,

$$\forall x \phi x \supset \phi y, \tag{2.15}$$

is preserved; its counterpart

$$\Box(\forall x \phi x \supset \phi y) \tag{2.16}$$

is not. For consider the following model:

$$
\begin{array}{llll}
w_1 & \phi u_1 \quad \phi u_2 & D_{w_1} = \{u_1, u_2\} \\
{\scriptstyle R(w_1,w_2)}\Big\downarrow & & \\
w_2 & \phi u_1 \quad \neg\phi u_2 & D_{w_2} = \{u_1\}.
\end{array}
$$

Given an assignment μ for which $\mu(y) = u_2$, (2.15) would be false in w_2, thereby making (2.16) false in w_1 for μ. As on the other hand (2.15) is true in w_1 for μ, the rule of necessitation fails.[6] (Hughes and Cresswell, 1996, p. 290)

Now, of course there is always a choice to be made whether intentional objects should be used to express the situation where something that exists in n-fold in one world exists in m-fold in another world (with $n \neq m$). The obvious alternative is to use regular objects and to flag some of these via predicates as not existing in some of the worlds. We believe that what should decide the question whether a and b in w_2 are counterparts of a in w_1 are, quite simply, the following formal criteria:

(1) If $R(w_1, w_2)$ then $w_1 \vDash \Box\phi a$ should entail $w_2 \vDash \phi a, \phi b$;

(2) If $R(w_2, w_1)$ then $w_2 \vDash \Box\phi a$ should entail $w_1 \vDash \phi a, \phi b$. Similarly, $w_2 \vDash \Box\phi b$ should entail $w_1 \vDash \phi a, \phi b$; and

(3) In w_1, it should be incoherent to state $P(a) \wedge \neg P(b)$ for all predicates P.

In our logic we get (2) and (3) for free because on our account, if a and b are counterparts in w_2 of a single object a in w_1, it should be the case that $w_1 \vDash a = b$. Criterion

---

[6]Note that V yields values for any tuple $\langle \phi x_1 \ldots x_n, w \rangle$ with ϕ a predicate, $w \in W$, and $x_1, \ldots, x_n \in D$—even some of these variables are not in the domain of the evaluation world. This may seem odd but the purpose of local domains D_w is in fact merely to restrict the range of objects that variables *quantify* over. (Garson, 2008, § 13) One might object that formulas containing some variable x to which $\mu(x)$ assigns an object not in the domain of the evaluation world, should have no value or be undefined. It turns out, however, that such modifications do not change the validity of well formed formulas in modal predicate logic. (Hughes and Cresswell, 1996, pp. 278ff) Hughes and Cresswell also show how certain other objections—such as insisting that variables are assigned only objects from the evaluation world and worlds recursively accessible from the evaluation world—ultimately fail (1996, pp. 290ff).

(1) poses more of a challenge. Hughes and Cresswell mention that replacing each formula $\Box\phi x$ by

$$\forall y(x = y \supset \Box\phi y) \tag{2.17}$$

allows one to express counterpart theory using intentional objects (1996, pp. 357–358). True, any implementation of counterpart theory in modal predicate logic is bound to presuppose a symmetrical counterpart relation (ibid.)—that is, it presumes that if a and b are counterparts in w_1 of a in w_2 then a is a counterpart in w_2 of a and b in w_1—but this shouldn't be a problem for what we want to do. On a semantic level the behavior of substituting (2.17) for $\Box\phi x$ can be directly attained by defining '$\Box$' as follows:

$$V_\mu(\Box\alpha, w) = 1 \iff V_\rho(\alpha, w') = 1 \text{ for every } w' \text{ such that } R(w, w')$$
$$\text{and every assignment } \rho \text{ such that for every free } x \text{ in } \alpha,$$
$$\text{there is a } y \text{ such that } \mu(x)(w) = \mu(y)(w),$$
$$\text{and 0 otherwise.}$$

(Cf. Hughes and Cresswell, 1996, pp. 354, 357–358)

In what follows we will, for simplicity's sake, assume intentional objects are implemented using a function I (as described above). We will also assume a fixed domain is used. As we have seen, however, there is no strict requirement to implement our logic like this. Indeed, because Kaplan rejects the rule of necessitation one of the most common arguments against world-relative domains is not applicable here. Furthermore, some readers may find that there are good philosophical reasons for using world-relative domains or alternative implementations of intentional objects.

2.3. RECONSTRUCTING THOUGHT

As discussed above, Kaplan makes a sharp distinction between the objects of thought and their cognitive significance. This is unfortunate because it leaves us without tools to analyze reference failures. In the remainder of this section we intend to develop a tool that *can* model such errors. Consider the following conversation:

Rachel : Do you remember my friend from French class?

Frank : I think so. Psychologist, blond hair, blue eyes.

Rachel : No, you're probably thinking of Caroline from cooking class. Macy
from French class is a journalist and brunette with gray eyes.

In this dialog, Rachel doesn't merely point out that Frank is mistaken about her friend's chosen profession and the color of her hair and eyes. She also makes the diagnosis that Frank was mistaking Caroline for Macy. Here's a basic model of this situation:

$$w_t \qquad \phi[\![\imath x \sigma x]\!] \quad \neg\psi[\![\imath x \sigma x]\!] \qquad \langle [\![\imath x \sigma x]\!], m \rangle$$

$R_\mathsf{C}(w_r,w_t) \Big\uparrow$

$$w_r \qquad \phi[\![\imath x \sigma x]\!] \quad \neg\psi[\![\imath x \sigma x]\!] \qquad \langle [\![\imath x \sigma x]\!], m \rangle$$

$R_\mathsf{C}(w_f,w_r) \Big\uparrow$

$$w_f \qquad \neg\phi[\![\imath x \sigma x]\!] \quad \psi[\![\imath x \sigma x]\!] \qquad \langle [\![\imath x \sigma x]\!], c \rangle.$$

In this model, w_t represents the world as it actually is, while w_r and w_f stand for the worlds as Rachel and Frank tried to reconstruct it. Additionally, ϕ is the predicate 'a journalist and brunette with gray eyes,' ψ means 'a psychologist with blond hair and blue eyes,' and σ tests for 'my friend from French class.' I use '$[\![\ldots]\!]$' for expressions that were evaluated in an earlier context; the result of these evaluations are listed in the rightmost column. As such we see that $\imath x \sigma x$, when stated by Rachel, refers to Macy (object m), is understood to refer to Macy by Rachel, but is mistakenly thought to refer to Caroline (object c) by Frank. Finally, the relation R_C indicates that w_f is an attempted reconstruction of w_r, which in turn is a reconstruction of the world of facts w_t.

Let's see what the semantics of '$\Box$' and '$\Diamond$' get us for my *reconstructivist interpretation of* R_C. Before we start, however, we'll rename these operators '$\mathsf{C}_\wedge$' and '$\mathsf{C}_\vee$' to indicate that we expect that we will no longer be looking at the usual interpretations—viz. "It is necessarily the case that ..." and "It is possible that ...." If some world w_1 has exactly one outgoing relationship R_C, say to w_2, then $\mathsf{C}_\wedge \alpha \iff \mathsf{C}_\vee \alpha$, for if α is true in w_2 then both $\mathsf{C}_\wedge \alpha$ and $\mathsf{C}_\vee \alpha$, and if α is false in w_2 then neither $\mathsf{C}_\wedge \alpha$ nor $\mathsf{C}_\vee \alpha$. On the other hand, if a world w_3 has no outgoing relationships R_C, then for every α, $\mathsf{C}_\wedge \alpha$ (trivially, if there are no accessible worlds then every proposition is true in all accessible worlds), but for no β, $\mathsf{C}_\vee \beta$ (no proposition can be true in *some* accessible world for there are no accessible worlds). Considering also worlds that have two or more outgoing relationships R_C, wherein the 'all' and 'some' semantics for $\mathsf{C}_\wedge$ and $\mathsf{C}_\vee$ are intuitively straightforward, we might attempt to formulate an interpretation for these operators.

The case of multiple outgoing relationships of R_C is perhaps the more interesting one, for if one attempts to reconstruct two accounts of the world, one is bound to be trying to reconcile incompatible views. For instance, consider the police officer who registers conflicting reports of eye witnesses, of whom the first witness says the robbers drove away in a blue SUV and the second witness says they escaped in a green SUV. Here it is a

true reconstruction that the robbers had an SUV (α) because both witnesses report α, thus $\mathsf{C}_\wedge\alpha$. On the other hand, we'd have *an inference to some explanation*—to take a cue from Bas van Fraassen—should a detective conclude (say) that the thieves drove a blue vehicle. An inference to some explanation α can be written down in our logic as $\mathsf{C}_\vee\alpha$.

Even though reconstructions of almost any two accounts are, *viewed in terms of worlds*, sure to entail a reconciliation of incompatible beliefs, reconstructions of two accounts needn't be similarly problematic if we restrict our view to certain *situations* within those worlds. Let us first, however, define situations. Semantically, we define a situation s as a function from n-place predicates to sets of n-tuples of the form $\langle d_1, \ldots, d_n \rangle$ with $d_1, \ldots, d_n \in D$. Situations defined this way are very similar to V, which is a function from n-place predicates to sets of $n+1$-tuples of the form $\langle d_1, \ldots, d_n, w \rangle$ with $d_1, \ldots, d_n \in D$ and $w \in W$. (Cf. Hughes and Cresswell, 1996, pp. 243, 333–335) Hence it should not come as a surprise that s contains propositions $\phi x_1 \ldots x_n$, and that the set of these propositions is a subset of the set of propositions to which V assigns truth values for each world.

The kind of situation we were hinting at earlier is this:

Definition 2.1. If in world w, for all propositions α belonging to situation s, $\alpha \in s \iff \mathsf{C}_\wedge\alpha \in w$ then we call s *fully reconstructed in w*.

To wit, all witnesses might just as well have seen the suspect drive away in a red SUV. If their beliefs about propositions unrelated to the crime scene—such as beliefs concerning the moral fortitude of the president of the USA—differ then that's quite alright.

Whenever a situation s is not *fully reconstructed*, there will be one or more propositions α that hold in s but for which it is not the case that $\mathsf{C}_\wedge\alpha$. For such propositions it is informative to know whether or not $\mathsf{C}_\vee\alpha$ is the case. Additionally, the following definition may prove useful.

Definition 2.2. If in world w, there is a proposition α in situation s, for which it is *not* the case that $\alpha \in w \iff \mathsf{C}_\vee\alpha \in w$, I call w a *reinterpretation of s*.

Reinterpreted situations by definition contain propositions that are not part of the accounts on which they are based. Thus for such situations we know that we are not dealing with a 'naive' reconstruction that merely combines the most plausible (for any definition of plausible) elements of each account. Rather, we would be dealing with a reconstruction that is misinformed or that second-guesses its supporting accounts. Such a revision could be introduced on the basis of theoretical background information—facts that are not in s—incompatible with the supporting accounts. For instance, suppose we believed that all supposed magic is actually merely clever exploitation of nature. On this

assumption, if we were to reconstruct someone's account of a wizard performing a miracle, we might reconstruct the event as the performance of an illusion.

2.3.1. Are Reconstructions Reflexive, Symmetric, or Transitive?

An interpretation of *many worlds semantics* in terms of reconstructions of accounts must not impose reflexivity on R_C. It is indeed intuitively felt that a reconstruction cannot very well be a reconstruction of itself. A more rigid analysis of the problem with reflexivity begins at the observation is that $\mathsf{C}_\wedge \alpha$ means that α is a true reconstruction. If $\mathsf{C}_\wedge \alpha \supset \alpha$ were valid (which it would be if R_C were reflexive) then all true reconstructions would be actual. This is undesirable because we want our interpretation to model *fallible* reconstructions—more clumsily we might say that we want to be able to model any *attempt* at a reconstruction. Modeling only those reconstructions w for which $\mathsf{C}_\wedge \alpha \in w \implies \alpha \in w$ is not a reasonable alternative. For consider any reconstruction w_r of the world of facts w_f (and of no other reconstruction). Here it so that for any $\alpha \in w_f \implies \mathsf{C}_\wedge \alpha \in w_r$. But if $R_\mathsf{C}(w_r, w_r)$ then also $\alpha \in w_f \implies \alpha \in w_r$. Thus everything that is the case in the factual world would also be the case in any reconstruction of it—including those propositions that are not part of any relevant situations. Clearly such reconstructions cannot exist on practical grounds and so R_C should be an anti-symmetric relation.

Symmetry does not make sense for R_C either. Suppose that w_1 is a reconstruction of w_2. Intuitively this relationship requires that w_1 was constructed based on observations of w_2, and so w_2 must have existed before w_1. Conversely, if w_2 were a reconstruction of w_1 then w_1 would have had to have existed before w_2. Obviously both relationships cannot exist at the same time and so not only must R_C not be symmetric, it must in fact be anti-symmetric on our interpretation. Indeed, R_C must not admit cycles of any kind because this would imply that the time dimension in which the different reconstructions were created, contained loops.

Finally, it might intuitively seem that it would make sense for R_C to be transitive some of the time. For if someone gives us his account of the factual world then sometimes we will want to reconstruct the factual world and other times we will want to reconstruct the witnesses's *personal account* of the facts. Reconsider the example of the police officer interviewing an eye witness. The officer should want to write down the witness's statements minding only what the witnesses tell him. A police detective reading the officer's report, on the other hand, should also be interested in what actually happened. Of course these modalities needn't yield the same thing. Specifically, it is possible that $\mathsf{C}_\wedge \alpha$ is true

when $C_\wedge C_\wedge \alpha$ is false, or *vice versa*. Now, of course, often you'll hear a report about this or that 'fact' α for which it is all but impossible to track down the source. What you're interested in is if α is true in the world of facts—i.e. if $\alpha \in w_f$. In our logic, so far, the only way to find out if $\alpha \in w_f$ is by checking the value for $C_\wedge \ldots C_\wedge \alpha$. But the problem is that you don't always know how many '$C_\wedge$' operators you need to arrive at the first world. Now, surely this problem could be solved by making R_C a transitive relation or by extending our logic with an additional operator '$C_\wedge^\rightarrow$' that is a transitive variant of '$C_\wedge$.' The semantics for such an operator can be added as follows:

$$V(C_\wedge^\rightarrow \alpha, w_1) = 1 \iff V(\alpha, w_2) = 1 \text{ and } V(C_\wedge^\rightarrow \alpha, w_2)$$
$$\text{for every } w_2 \text{ such that } R_C(w_1, w_2). \qquad (VC_\wedge^\rightarrow)$$

(Cf. Hughes and Cresswell, 1996, p. 243) What this operator says is that α has been communicated from *some* first world down to the actual world without mutations in intermediate worlds. This doesn't tell us all that much, however, because it doesn't tell us which world exactly is this first world. For instance, if α stood for "In a galaxy, far far away, an evil empire wrecks havoc" then $C_\wedge^\rightarrow \alpha$ might be true if you overheard a conversation about *Star Wars*, even though α wouldn't be a fact. It gets worse. For any made up claim β it is always trivially the case that $C_\wedge \beta$ when β is originally reported because made up claims, not being reconstructions of any other world, do not have access to any worlds over R_C. Consequently the usefulness of '$C_\wedge^\rightarrow$' is greatly diminished. Particularly, there is no rationale for making '$C_\wedge$' transitive by default and so we shouldn't want to replace R_C by a transitive counterpart.

2.3.2. Empirical Basis and Intended Target of Reconstruction

We have argued that R_C should neither be reflexive, symmetric, nor transitive. Indeed it should be anti-reflexive and acyclic. However, in the process of explaining why R_C shouldn't be transitive, we also demonstrated the $C_\wedge$-operator doesn't *always* quite get us what we want. The problem, we believe, is that there are multiple interpretations for R_C. The interpretation we have been following so far is that $R_C(w_0, w_1), \ldots, R_C(w_0, w_n)$ indicates that w_0 tries to conform to all w_i (with $1 \le i \le n$) as much as possible by way of copying the data of all w_i. This interpretation is a causal one. It stipulates that $w_1, \ldots, w_n$ are the empirical basis for w_0. This relation can be formulated in English as follows: 'To reconstruct from $w_1, \ldots, w_n$ (as w_0).' A second interpretation makes R_C a relation that reconstructs *something*. Suppose w_f is the worlds of facts in which a bank

robbery took place, $w_{w_1}, \ldots, w_{w_n}$ are eye witness accounts to the robbery, and w_r is a police report of the crime. The relation between these worlds can be put in English like this: 'To reconstruct w_f from $w_{w_1}, \ldots, w_{w_n}$ as w_r.' Here w_f is the world that the police *intends* to reconstruct.

Depending on circumstances we might be interested in whether a reconstruction successfully captures its empirical basis, if it gets the facts that it is meant to reconstruct right, or both. Consider again the example of a crime scene. Here the police officer interviewing the witnesses should write down a report w_r for each witness i so that for every statement α in a situation s, $\alpha \in w_r \iff \alpha \in w_{a_i}$ (w_{a_i} being the account as witness i actually told it). A detective, on the other hand, does not (in the first place) care so much about what witnesses say happened as about what has *truly* happened. Thus the detective's reconstruction should ideally meet the constraint $\alpha \in w_r \iff \alpha \in w_f$ (with w_r being the detective's reconstruction and w_f being the bare facts). Sometimes we want to know if some proposition α is both the case in our empirical basis and in the world we intend to reconstruct—for instance, when we want to test the reliability of our empirical base.

Consequently, what we seem to need is an additional operator D that has the same semantics as $\mathsf{C}_\wedge$ except that it uses accessibility relation R_D instead of R_C. Like R_C, R_D should be neither reflexive, symmetric, nor transitive. Like R_C, it should also be anti-reflexive and acyclic. We further specify that for each world $w_r \in W$ there is at most one world $w_f \in W$ so that $R_\mathsf{D}(w_r, w_f)$ (as any two worlds are bound to differ, it would seem that one cannot sincerely intend to capture them both in a single reconstruction). By these semantics $\mathsf{D}\alpha$ will be true in w_r for all α that are not false in a world w_f so that $R_\mathsf{D}(w_r, w_f)$. It is worth special mentioning that $\mathsf{D}\alpha$ will be true for literally every α if w_r does not intend to restruct any world w_f (indicating that there are no prior constraints to what is truthful).

2.3.3. Do Reconstructions Model Belief States?

Reconstructions, as I have used the term in this paper, do not model belief states as such. Firstly, not all propositions in a reconstruction are likely to be beliefs and neither is there a formal constraint that every belief must be represented in some reconstruction. Secondly, reconstructions contain not only cognitive content but also meta-content about thoughts.

Reconstructions are modeled by worlds. This implies that, for every n-place predicate ϕ and all permutations of objects $x_1, \ldots, x_n$, they have a notion whether $\phi x_1 \ldots x_n$ is

true. Now, clearly agents do not have beliefs about all predicates and objects and so reconstructions cannot be belief states. On the other hand, for every reconstruction it should be possible to define a nonempty situation s such that every proposition in s corresponds to a belief. For instance, if w_a is a reconstruction by agent a of agent b's account w_b of how the world is (w_t) then there is a situation s so that

$$\alpha \in s \implies (\alpha \in w_a \ \& \ \mathsf{B}_a\mathsf{B}_b\alpha \in w_t)$$

and, similarly,

$$\alpha \in s \implies (\alpha \in w_b \ \& \ \mathsf{B}_b\alpha \in w_t)$$

where $\mathsf{B}_i\alpha$ stands for 'agent i believes that α.' Notice how meta-knowledge about what w_a reconstructs is needed to translate α into $\mathsf{B}_a\mathsf{B}_b\alpha$. Also note that this translation only works because the source world here is w_t—the actual world of facts. Had agent b explained the *Star Wars* universe (w_s) to agent a instead, things would have been different: It would have neither been the case that $\mathsf{B}_a\mathsf{B}_b\alpha \in w_s$ (the *Star Wars* universe doesn't know about agents a and b) nor that $\mathsf{B}_a\mathsf{B}_b\alpha \in w_t$ (both agents know that *Star Wars* isn't real). On a third scenario wherein b is explaining the *Star Wars* universe to a, but a thinks b is speaking about events that actually took place long ago, it would be the case that $\mathsf{B}_a\mathsf{B}_b\alpha \in w_t$ yet— presuming b realizes that *Star Wars* is fiction—$\mathsf{B}_b\alpha \notin w_t$. We thus find that when mapping a reconstruction into belief statements, we need to know what the agent responsible for the reconstruction *believes* it is reconstructing. It follows that an important difference between the B-operator and reconstructions is that the former assumes it is transparent to the agent what it has beliefs about, whereas reconstructions allow us to model confusion on the part of the agent about what it is reconstructing. Put differently, B-propositions concern belief content only, whereas models of reconstructions also record part of how beliefs come about. Lastly, observe that translating reconstructions of fictional worlds into belief propositions involves creating new predicates or objects. After all, should Oscar reconstruct Orwell's *1984* universe and write down that the government is spying on its citizens, this doesn't mean Oscar believes his government is spying on him. Instead it means that he believes that 'the *1984* government' is spying on 'its *1984* citizens.' Or we could say he believes that the government is in an 'is represented in *1984* as spying on' relationship to its citizens.

Our formal system does not require that every belief is represented in some reconstruction. Such a constraint *could* be added, but this may be undesirable. For consider

that in epistemic logic it is usually held that $B_a\alpha \supset B_aB_a\alpha$. That is, if agent a believes that α then it will also believe that it believes that α. Thus, the accessibility relation for B must be transitive. It was suggested above that the correct translation of $B_a\alpha$ is that a has a reconstruction of the world of facts and that in this reconstruction α is the case. Similarly, $B_aB_b\alpha$ should, in our terminology, mean that a has a reconstruction of b's reconstruction of the world of facts, and that in a's reconstruction it is the case that α. But then for every reconstruction w_a for which there is a reconstruction w_b so that $R_D(w_a, w_b)$—that is, w_a reconstructs *something*—there must be another reconstruction w_c so that $R_D(w_c, w_a)$. As we already decided that R_D is an acyclic relation, this would mean that as soon as a world w_a is made accessible from another reconstruction w_b, an infinite chain of reconstructions must be admitted. We probably want to keep the number of reconstructions finite though—this way we can restrict the moniker 'reconstruction' for those reconstructions that someone actually bothered to make. In other words, R_D mustn't be serial and therefore we must admit that not all beliefs, but rather only those beliefs that are thought, are reconstructions.

2.3.4. Redefining Context of Utterance

K. Romdenh-Romluc (2006) brings up some interesting counterexamples to Kaplan's theory of indexicals. For starters, suppose you were reading a book on the life of the Buddha. Somewhere at the start of the book it reads "Now Siddharta leaves the palace" (ibid.). Kaplan's theory would seem to predict that the indexical 'now' here refers to the moment the book's author wrote down that sentence. But obviously it instead refers to a moment in the life of Siddharta. What gives? It is our view that the meaning of pure indexicals such as 'I,' 'now,' and 'here' are embedded in reconstructions. Additionally we believe that 'stories' are a reconstruction genre that is characterized by the modification of terms such as 'here' and 'now' to closely fit the situation described. For example, historical stories will typically rewrite the terms 'here' and 'now'—but, contrary to most fictional stories, not 'I'—to fit the location and time that is being described. Thus, whereas in the world of facts that we inhabit most of the time, the word 'here' means something along the lines of 'the location of the agent using the word "here,"' in a biography it might mean 'the location of our subject at the described time.'

Let us pause for a caveat: The question what reconstruction an indexical is to be evaluated in remains unanswered. What we do claim, however, is that the hardest part of solving this question consists of solving the question what reconstruction *any* term is to be evaluated in. Consider the sentence "Now, finally, Siddharta leaves the palace—

of course, now tourists from all over the world come to see Siddharta's home" (ibid.).
Here the first occurrence of 'now' should be evaluated in the same reconstruction as the
phrase "Siddharta leaves the palace." The second mention of 'now' belongs in the same
reconstruction as "[T]ourists from all over the world come to see Siddharta's home." To
wit, the first reconstruction—the story about Siddharta—is a reconstruction about the
world of facts that utilizes the pure indexical 'now' as a storytelling device; the second
reconstruction is a regular reconstruction of the same world of facts, but doesn't take
recourse to such tricks. We do not have a solution to this problem of pragmatics, but
it should be interesting enough by itself that when a proposition uses indexicals, the
characters of those indexicals are determined by the reconstruction that the proposition
belongs to. In any case, reconstructions here prove germane to the topic of indexicals
even beyond issues concerning cognitive content of thought.

What I have said may seem to fly in the face of everything Kaplan discovered. Gone
is the distinction between the context of utterance and the circumstances of evaluation.
Well, not exactly. Recall that Kaplan introduced the distinction between the context of
utterance and the circumstances of evaluation to make indexicals work in counterfactuals
of the type "I could have not been here right now." So let us investigate counterfactuals a
bit closer. We could think of counterfactuals as 'reconstructions' that intentionally twist
facts. We could also add that the genre of counterfactuals has the characteristic that
the pure indexicals have characters such as 'the person *actually* uttering the word "I."'
So it turns out that the trick is to define an 'actually' operator that, for counterfactuals,
evaluates propositions in the world that is being parodied.[7] In all other genres it evaluates
propositions in the current world.

Hughes and Cresswell describe an implementation of an 'actually' operator using dou-
ble indexing (1996, pp. 350–352). In our logic such an implementation would require
firstly that we yet again modify the evaluation procedure for atomic formulas:

$$V_\mu(\phi x_1, \ldots, x_n, \langle w_1, w_2 \rangle) = 1 \iff \langle \mu(x_1)(w_1), \ldots, \mu(x_n)(w_1), w_1 \rangle \in V(\phi).$$

The tuple $\langle w_1, w_2 \rangle$ in the left side of this formula is a tuple of the world in which $\phi x_1 \ldots x_n$
is being evaluated (here w_1) and the 'actual' world (here w_2). Next we would be able to
introduce the new operator 'A:'

$$V_\mu(A\alpha, \langle w_1, w_2 \rangle) = 1 \iff \langle \alpha, \langle w_2, w_2 \rangle \rangle \in V(\phi).$$

---

[7]Kaplan confirms that his distinction between the contexts of use and evaluation can be formalized
using an 'actually' operator (1989*b*, pp. 594–596).

That is, evaluating $A\alpha$ in any world is equivalent to evaluating α in the 'actual' world. Finally, in all other semantic rules we'd have to substitute $V(\alpha, \langle w_x, w_3 \rangle)$ for every $V(\alpha, w_x)$.

Double indexing does not quite get us what we want. We'd like to be able to mark some reconstructions with the metafact that they are counterfactual worlds. This should make it unnecessary to manually specify the 'actual' world every time a formula is evaluated. In other words, we want the world that is 'actual' for some counterfactual world to be fixed, whereas if we used double indexing it would be dynamically specified at evaluation time. The solution is to introduce one more accessibility relation R_A, which for every world w_1 designates exactly one other world so that $R_A(w_1, w_2)$. The idea is that propositions in w_1 that have the form $A\alpha$ will be evaluated as α in w_2. In most reconstruction genres it will be the case that $w_1 = w_2$. In the counterfactual genre w_2 will be accessible from w_1 over a chain of worlds connected by R_C. For most counterfactual 'reconstructions' w_2 should be removed from w_1 by just a single step; however, for counterfactual reconstructions of counterfactual reconstructions, the distance will be at least two worlds; and so on. We can now define the 'actually' operator as follows:

$$V_\mu(A\alpha, w_1) = 1 \iff V_\mu(\alpha, w_2), \text{ with } w_2 \text{ the one world so that } R_A(w_1, w_2).$$

If modal operators were introduced, any possible world w_p accessible from a reconstruction w_r via any number of nested '$\Box$' and '$\Diamond$' operators would define a relationship $R_A(w_p, w_a)$ so that w_a also fulfills $R_A(w_r, w_a)$.

Double indexing invalidates the rule of necessitation, which Kaplan thinks is a good thing because he wants the statement "I'm here right now" to be a tautology (because it cannot be uttered falsely) but not a necessary truth. Our logic of reconstructions has a similar property, although it is less powerful. In many reconstruction genres the sentence "I am here right now" will always be true. In counterfactual situations, however, it can also be false. Thus the more specific property of being tautological *within certain reconstruction genres* is here substituted for the property of being tautological. Of course this is a much weaker property than what Kaplan's logic offers. However, ignoring counterfactuals for a moment, the fact that we allow for reconstructions in which "I'm here right now" is not always true (e.g. historical texts) is intentional. General tautology simply is too strong a property on our analysis. If we only take in account (1) reconstructions in which "I am here right now" is true and (2) counterfactuals then we find that "I am here right now" is true in just those worlds w for which $R_A(w, w)$. This is interesting because Kaplan calls formulas valid if are true in the *circumstances of their context* (cf. Kaplan, 1989a, XIX), which in the terminology of doubly indexed logics means that formulas are

valid if they are true when the evaluation and 'actual' worlds are the same.

It may seem strange that the 'actually' operator does not always cause propositions to be evaluated in the world of facts. However, our concern here is with dthat terms, such as dthat[the guy behind the three], that have to be evaluated within, say, a—possibly incorrect or even incoherent—retelling of a fictional story. When we need to know if the retelling was correct, we can still use $\mathsf{C}_\wedge$, $\mathsf{C}_\wedge^\rightarrow$, $\mathsf{C}_\vee$, and D.

Granted, sometimes we really do want to leave the story we are telling for a moment and point to an object in the factual world—like when someone is on the witness stand and points to the perpetrator. But then we might also point to two different comic books—which are reconstructions by themselves—and say 'this story's superhero' could beat up 'that story's superhero.' So it's not just that we'd like an additional operator to talk about the factual world; rather, we want an operator that allows us to evaluate singular terms and propositions in arbitrary reconstructions of our choosing. Such an operator would make for an interesting extension to our language because ideally we should be able to reason about all aspects of reconstructions in our language. Time constraints, however, prevent us from exploring this topic any further.

2.3.5. Modelling the Different Types of Reference Failure

To reiterate, there are three noteworthy types of reference failure with respect to indexicals:

(1) The first kind of reference failure is when one holds some indexical to refer to object a when it actually references some other object b;

(2) The second error entails thinking an indexical refers to some object c when c is but an artifact of the reconstruction and does, in fact, not exist;

(3) The third mistake involves believing that two indexicals refer to two distinct objects d and e when they actually refer to the same object.

Given the logic outlined above, situation (1) can be modeled as follows:

$$w_f \qquad \phi a, \neg \phi b \models \neg \phi [\![\imath x \sigma x]\!] \qquad \langle [\![\imath x \sigma x]\!], b \rangle$$

$$R_\mathsf{C}(w_r, w_f) \Big\uparrow$$

$$w_r \qquad \phi a, \neg \phi b \models \phi [\![\imath x \sigma x]\!] \qquad \langle [\![\imath x \sigma x]\!], a \rangle.$$

In the rightmost column we see that w_r and w_f disagree on which object meets $\imath x \sigma x$ in the context of use. In the middle column we see that, presuming *ceteris paribus*, w_r

and w_f agree on the properties of objects in the circumstances of evaluation. The model demonstrates that equal circumstances of evaluation can lead to contradictory conclusions if the perception of the context of use differs.

Because we are working with fixed domains and an existence predicate, (2) is really just a special case of (1):

$$w_f \qquad\qquad \neg E(c), E(z), \phi c, \neg \phi z \models \neg \phi [\![\imath x \sigma x]\!] \qquad\qquad \langle [\![\imath x \sigma x]\!], z \rangle$$

$$R_{\mathsf{C}}(w_r, w_f) \Big\uparrow$$

$$w_r \qquad\qquad E(c), E(z), \phi c, \neg \phi z \models \phi [\![\imath x \sigma x]\!] \qquad\qquad \langle [\![\imath x \sigma x]\!], c \rangle.$$

Finally, scenario (3) gives us the following results:

$$w_f \qquad \phi d, \phi e \models \phi [\![\imath x \psi x]\!], \phi [\![\imath y \sigma y]\!] \qquad \langle [\![\imath x \psi x]\!], a/b \rangle, \langle [\![\imath y \sigma y]\!], d/e \rangle$$

$$R_{\mathsf{C}}(w_r, w_f) \Big\uparrow$$

$$w_r \qquad \phi d, \neg \phi e \models \phi [\![\imath x \psi x]\!], \neg \phi [\![\imath y \sigma y]\!] \qquad \langle [\![\imath x \psi x]\!], d \rangle, \langle [\![\imath y \sigma y]\!], e \rangle.$$

By premise of (3) there is a single object in w_f that has two counterparts in w_r. These two counterparts are assigned to d and e as intentional objects with different extensions in D. When evaluated in w_f these same intentional objects d and e result in a single shared extension in D. Now, what happens in w_f is that $[\![\imath x \psi x]\!]$ and $[\![\imath y \sigma y]\!]$ yield intentional objects that are indistinguishable from d and e. In w_r, on the other hand, they yield intentional objects that are distinguishable from each other—that is, $[\![\imath x \psi x]\!] \neq [\![\imath y \sigma y]\!]$.

So far my analysis of (3) does not explain why we need intentional objects. It might seem that merely making $[\![\imath x \psi x]\!]$ and $[\![\imath y \sigma y]\!]$ refer to the same object d—thereby ignoring e—would get us the same results. Intentional objects prove useful, though, once we start checking if this and that proposition are true reconstructions. For instance, observe that whereas in w_r it is true that $\phi [\![\imath x \psi x]\!]$ but false that $\phi [\![\imath y \sigma y]\!]$, it is both true in w_r that $\mathsf{C}_\wedge \phi [\![\imath x \psi x]\!]$ and $\mathsf{C}_\wedge \phi [\![\imath y \sigma y]\!]$. Similarly, whereas $[\![\imath x \psi x]\!] \neq [\![\imath y \sigma y]\!]$ in w_r, it turns out that in the same world $\mathsf{C}_\wedge [\![\imath x \psi x]\!] = [\![\imath y \sigma y]\!]$. Recall that on our interpretation this signifies that construction w_r is faulty where it reconstructs $[\![\imath x \psi x]\!]$ and $[\![\imath y \sigma y]\!]$ as demonstrating different objects.

There actually is a fourth kind of reference failure. We didn't mention it earlier, however, because in a way it is an artifact of our logic of reconstructions. What can happen is that an agent is mistaken about what worlds it is reconstructing. Specifically, for a reconstruction w_a there may be a world w_b so that either of $R_{\mathsf{C}}(w_a, w_b)$ and $R_{\mathsf{D}}(w_a, w_b)$ holds, but not both. Or this situation may occur for any w_c linked to from w_a by a chain of worlds over R_{C} or R_{D} (or both). For instance, a careless agent may provide a

detailed description of Luke Skywalker, yet believe to be explaining what *Star Trek* is about. This fourth type of reference failure may explain an interesting paradox discussed by Romdenh-Romluc (2006, p. 266). Suppose Amanda works for a radio station on which she has a history show. She also sometimes does the news. On one particular occasion she is supposed to read the news, but Amanda mistakenly brought her notes for the next episode of her history show—a continuation of the previous episode. It is announced that Amanda will read the news, but then she says

$$\text{Now the French are invading England!} \qquad (2.18)$$

(Ibid.) On our logic both Amanda and the listeners will be reconstructing a twentieth century historical event. Whereas Amanda also believes to be reconstructing this event, however, her listeners think she is reporting on a current event. In other words, in the reconstruction that Amanda makes of her own actions it is the case that $\mathsf{D}(2.18)$, whereas in the reconstructions that her listeners make of what they hear, $\neg\mathsf{D}(2.18)$ holds. Now, we take Romdenh-Romluc's example as a challenge to explain why (2.18) is *objectively* false. Decomposing the event may be key to explaining this. We define three situations:

(1) s_1 contains propositions that assert that it was announced that Amanda would read the news (e.g. propositions such as "It was announced that Amanda would read the news," rather than "Amanda will read the news");

(2) s_2 contains propositions that assert that Amanda would continue her history program, following up on the previous episode about the Norman Conquest, and so on; and

(3) s_3 contains the propositions asserted by Amanda herself.

Notice that s_1 and s_2 are competing accounts of the stage setting for s_3. It is our view that this stage setting, combined with various aspects of pragmatics, determine what exactly, in an objective sense, s_3 is supposed to reconstruct. As all propositions in s_1 are true in the factual world that they reconstruct, but at least some of the propositions in s_2 are false in the factual world, we know that Amanda's reconstruction of the event is liable to be rejected by those who learn of it. In analogy to R_C and R_D we could introduce another relation R_E that would mark what world is *objectively* reconstructed. We could then also add an operator E that would tell us which of the propositions in s_1 and s_2 were correctly reconstructed. From an epistemological point of view, however, we do not need such an operator. Knowing that (some of) the propositions in s_2 are false, we know enough to

cast *doubt* on reconstructions of the present geopolitical world that hold the propositions in s_3 to be true. We know then that such reconstructions cannot represent knowledge, even if we weren't in an epistemic situation to know that the propositions in s_3 were false. In the next chapter we will discuss epistemological properties of reconstructions in more detail by relating our logic of reconstructions to epistemological contextualism.

3. Epistemological Contextualism

Contemporary contextualist epistemologists are divided foremost by the question: What determines a context? One influential answer has it that the person assessing a claim to knowledge determines the context in which that claim is to be evaluated. This view is known as attributer contextualism and its most popular version is *conversational contextualism* (CC)—as advocated by Cohen, Lewis, and DeRose. According to this loose alliance of theorists, attributer contextualism needs to be modified as follows: It is not so much the person *assessing* a claim to knowledge who determines the context—as the attributer contextualist thesis holds—but instead it is the person making an utterance of the form "S knows that p" who fixes the situation in which this epistemological proposition is to be judged. Orthogonal to attributer contextualism is subject contextualism[1], which traces epistemological contexts back to the subject S who may or may not know that p. Two proponents of this approach are Williams and Barke. (Brendel and Jäger, 2004)

In this section we will go through some of the arguments and paradoxes that have shaped the theoretical landscape sketched above. Williams remarked that overcoming skepticism is a defining aim of epistemology (2001, pp. 2–5). In any case, neutralizing skepticism does appear to be a central goal of epistemological contextualism. (E.g. see Brendel and Jäger, 2004, p. 147) Thus, it seems like a good idea to start our endeavor with a look at some skeptical problems and their contextualist remedies.

3.1. HOW TO DEFEAT SKEPTICISM

3.1.1. Local Skepticism

Local skepticism, which is perhaps the simplest of skeptical arguments, works by exploiting the mere possibility that a piece of evidence may be flawed, to demonstrate that one does

---

[1]Some authors prefer the moniker subject-sensitive invariantism for this type of contextualism.

not *know* a given fact (i.e. you're not certain of it). For instance, say you went to the zoo and reported back that you saw a zebra. A skeptic might reply that, for all you know, you might actually have seen a painted a mule instead. Therefore, the skeptic argues, you don't really *know* you saw a zebra. (Williams, 2001, pp. 73–75, Brendel and Jäger, 2004, p. 144)

Fred Dretske's solution to local skepticism comes down to defining knowledge as follows: S knows that p if and only if S is in an epistemic position from which he can rule out all relevant alternatives to p. (Alternatives to p are here understood as propositions that imply $\neg p$.) Now, for Dretske it is the epistemic situation of the subject that makes a given alternative relevant. So on his theory you truly might know you saw zebras at the zoo. However, you wouldn't have *known* you saw them—even if you did—should it be common knowledge that the zoo you visited had in the past tried to pass of mules for zebras. Not unless you had taken extra steps to verify the zebras were really what they appeared to be anyway. (Brendel and Jäger, 2004, p. 144)

Unlike in Dretske's theory, in CC the epistemic position of the subject does not determine the active context. Instead context changes are taken to be induced when the speaker's attention is drawn to an error-possibility. Changing contexts may here—in CC—be understood simply as the raising or lowering of standards for what counts as knowledge. Accordingly, a person who is, spoken from one context, said to know something, may turn out to not know that very same thing viewed from a different context (with higher standards). The approach to bind contexts to conversational dynamics is sometimes called indexicalism. The thinking leading to this name is that CC is a semantic theory stating that the phrase "S knows that p," is analogous to indexical assertions such as "X is flat." The semantics of 'knows' might then come to vary during the course of a conversation, much like the meaning of 'is flat' depends on the conversational context in which this predicate is used. We can now describe CC's solution to skepticism about zebras as follows: If "S knows that he saw a zebra" is true in one context, the same statement might yet be false in another context wherein the possibility of deceit has been mentioned. (Brendel and Jäger, 2004, § 3)

There are several problems with CC which have led us to regard this theory as inadequate. For starters, we find that some of the linguistic evidence gathered by Wayne Davis, makes CC implausible from a descriptive point of view (2004, p. 266). Consider a Luxembourger claiming the Netherlands are big and a Frenchman disagreeing. Depending on our own standards, we might agree with either fellow. But, more importantly, we might also exclaim: "They're both right!" Conversely, it would seem that two epistemologists disagreeing on whether or not they saw a zebra cannot possible *both* be right (if their dis-

pute is epistemological in nature). A second problem is the *paradox of epistemic laziness* (Brendel and Jäger, 2004, p. 150). In CC, mentioning error possibilities raises epistemic standards, and thereby subtracts from the set of known propositions. This is odd because this means that by merely considering skeptical arguments against what one knows, one destroys one's knowledge. Conversely, a less critical individual—speaking from an epistemological point of view—may end up 'knowing' more things precisely because he is not as critical. This is inconsistent with Frank Hofmann's remark that knowledge should be an *achievement* (ibid.). We will cite additional arguments against CC throughout subsequent sections.

Barke believes context changes cannot be explained solely by conversational dynamics. That's why her theory includes a new procedure for resolving skeptical attacks. The methods we apply when searching for answers to a certain question, Barke holds, work properly only given certain conditions. Therefore, when applying these methods, we implicitly seem to assume these conditions are *in fact* fulfilled. Now, a situation may arise wherein we come to question our assumptions. If this happens, we are faced with two alternatives: (1) Drop the assumption—possibly adopting a new method of inquiry that doesn't rely on said assumption—or (2) legitimate the assumption—perhaps after having conducted an investigation independent of current inquiries into it. It seems to us that this procedure is an important improvement on CC and so we will return to it shortly. But let us first see what the two remaining kinds of skepticism are about. (Barke, 2004, § 4)

3.1.2. Cartesian Skepticism

Cartesian skepticism, or the problem of *radical underdetermination*, is best understood as a generalization of local skepticism: The Cartesian skeptic holds that no evidence is ever strong enough to make some belief or other sufficiently probable so as to rule out all alternative explanations—bizarre as some of those alternatives may be. One well known argument for Cartesian skepticism claims that, for all we know, the external world might be an illusion and we might just be brains in vats. And from this—the argument continues—it follows that everything we think we know, including the common sense proposition that we have hands, might be false and is therefore not true knowledge. (Williams, 2001, § 6, Brendel and Jäger, 2004, p. 144)

Williams remarks that Cartesian skepticism is reducible to the observation that one may make standards for knowledge so tough so as to make it impossible for anything to be known (2001, pp. 75, 188, 195–196; 2004, § 5). But then, he argues, fallibilism—

the thesis that things we're justified to believe we know might be false—is nowadays commonly accepted anyway. So why give up our stakes in knowledge? We shouldn't, says Williams. We need only observe that the fashion in which we judge knowledge when we're practicing epistemology, is nontransferable to everyday life. That's because *epistemology proper* covers a subject—knowledge as such—different from the epistemological issues in our daily lives—which concern very specific claims to knowledge.

Dretske and Robert Nozick offer a different solution to Cartesian skepticism: They reject the *principle of epistemic closure* (PEC). This principle (in its basic form) states that if S knows that p, and knows that $p \supset q$, then S also knows that q. The argument that we may be brains in vats and therefore don't know that we have hands, as outlined above, employs this principle: It says that if having hands implies not being a brain in a vat and if we don't know that we're not brains in vats, then we don't know that we have hands. Reject the PEC and Cartesian skepticism dissolves. Not all authors think we should reject the PEC though. DeRose, for instance, finds the conjunction of not knowing we aren't brains in vats but still knowing we have hands, intuitively bizarre. (Brendel and Jäger, 2004, § 2)

As we reported earlier, one problem with CC is that it says the mere mentioning of a (skeptical) error possibility should raise knowledge attribution standards. We already noted that this property is undesirable from a normative point of view. From a descriptive point of view, we also doubt its predictive value. In fact, an example by Richard Feldman teaches us that this property *qua* prediction is just odd when considering Cartesian skepticism (Brendel and Jäger, 2004, p. 149). Suppose that you're in a conversation that is turning dull fast. In an attempt to change the topic to something more to your liking, you exclaim: "Well, at least I still know that I'm not a brain in a vat." CC seems to suggest that this utterance should invoke a context of higher standards, thereby making your statement contentious. What is more likely to happen, though, is that you'll be reminded that society at large doesn't consider such epistemological issues pressing at all. All needn't be lost for CC *per se*: Cohen constructed an *inattention thesis* and an *error theory* that account for this indifference by saying we sometimes simply fail to notice context changes. (Brendel and Jäger, 2004, p. 156) However, we agree with Davis that this explanation is *ad hoc* (2004, p. 266).

3.1.3. Agrippian Skepticism

Agrippian skepticism starts from the observation that people can disagree on anything. As such, we might plausibly argue that any claim that p, should first be understood as

"*S* believes that *p*." If, additionally, we want to say that *p* is *truly* the case, we have an obligation to offer evidence for *p*. However, as intrinsically all evidence for *p* will consist of claims to knowledge, we will also need to provide evidence for this evidence. Should we not wish to go on arguing forever, the skeptic remarks, we have but two options: Either we make a dogmatic presumption at some point—but that means giving up the project of founding our knowledge—or we reuse a claim we made earlier in our demonstration—but then our reasoning will have been viciously circular. We lose either way. (Williams, 2001, § 5)

Williams's cure for Agrippian skepticism is the *default and challenge model of justification* (DCMJ), which for its central tenet has it that epistemic subjects need not ground all of their knowledge claims (2001, pp. 148–150, 159–162). Williams's DCMJ only requires subjects to back up those knowledge ascriptions to which legitimate criticism—that is, criticism which casts plausible doubt on the reliability of the subject's sources of knowledge—was raised. This shift of the burden of proof stops vicious Agrippian regressions by introducing *default positions* for subjects to rely on until said positions are discredited. Finally, by Williams's *issue contextualism* (IC), the issue of whether or not a belief may be used as a default position is decided by the active context.

Hofmann remarks that Barke is in trouble if the reliability of sense perception is challenged (2004, p. 379). After all, every empirical method relies on the reliability of senses, yet philosophical thought is yet to yield satisfactory grounding of it. Agrippian skeptics can easily exploit this flaw in Barke's *epistemic contextualism* (EC) to demonstrate that empirical 'knowledge' is not the real thing—it's just how things *appear* to be. Here IC proves superior: Williams's DCMJ requires only that *legitimate* criticism of epistemic claims is rebutted. And where's the *prima facie* legitimacy in saying our eyes deceive us in everyday situations? Now, we feel it's warranted to say that sense perception is reliable in everyday situations but how do we know if sense perception is reliable when we're in an *unusual* situation? Indeed, how do we tell we're in an unusual situation in the first place? What we want to know is this: By what general procedure do we decide if a skeptical challenge is legitimate? Additionally, how do we contextualize knowledge of the appropriateness of challenges? To answer these questions we will need to take a closer look at how IC contextualizes the justification of knowledge claims.

3.2. A CLOSER LOOK AT ISSUE CONTEXTUALISM

Barke holds that the main thesis of epistemological contextualism is the proposition that every knowledge claim has a context, which in part determines if the claim is true or false

(2004, p. 353). She believes the two main tasks facing a contextualist are as follows:

(1) To give a plausible explanation of what constitutes a context; and

(2) To describe the causes and dynamics of context changes.

Although we shall indeed explain what contexts are made of and how they change, we have decided to not answer these questions separately. It is, we believe, best to not follow Barke's scheme, for Williams talks about contexts in a manner that is fundamentally different from Barke and the conversational contextualists. For most contextualists, events initiating context changes seem to be easily discernible as such. For instance, in CC, a context change might be triggered by the following question: "Have you considered you might have seen a painted mule?" Similarly, Barke thinks the ritual for forcing context changes consists of two parts: (1) This or that assumption must be called into question, and, subsequently, (2) the subject must fail to adequately defend the assumption. Previously we found ourselves encouraged by the simplicity of these models to visualize contexts as tags attached to subjects (Barke) or knowledge attributers (CC). The tags would contain only epistemic information on how to interpret knowledge claims. This information might be like a number, signifying the current level of scrutiny (CC again); or, alternatively, the tags might list assumptions made by the subject (Barke). Most of the time, we figured, epistemic parties would just read off information from the tags. This information, together with instructions encoded in epistemological papers and textbooks, could then be used to make epistemic assessments. Only on certain well defined and epistemically significant events someone—perhaps a neutral third party—would take gum and pencil, and update the tag (i.e. change the context).

The picture sketched above is too simple to describe IC. For starters, Williams does not speak of context changes. Instead he talks about changing the subject, and about raising standards. Williams also doesn't appear to think of contexts as tags or registers that keep track of a special—epistemic—kind of information. Rather, he posits that the epistemic status of knowledge claims is contingent on a total of five kinds of factors, which contextualize these claims. As a result, thinking of contexts as if they were some kind of entity—or data structure for that matter—is a bad approach to understanding IC.[2]

Let us first point out that there are two things that Williams wants to get *out* of contexts:

---

[2]Because we did not grasp this unique character of Williams's theory when writing our Bachelor's research paper, much of our previous criticism against Williams turns out to have been unwarranted. Furthermore, as a—somewhat indirect—consequence of this, much of our at the time supposed originality now seems to have been mere vanity. *Mea culpa!*

(1) Contexts should guard us from local and Agrippian skepticism by enabling us to dismiss all too remote error possibilities; and

(2) They should inform us when an argument goes off topic (like when the Caretesian skeptic starts discussing knowledge as such).

To realize purpose (1), Williams first embeds epistemic issues into a conversational context. It is here that default positions first turn up. Recall the skeptical challenge that perhaps we saw painted mules—not zebras—at the zoo. Should a skeptic point out this error possibility right away as we are telling about our visit to the zoo, we would be justified to dismiss his far fetched story. It would be a default position that what we perceived as zebras were, in fact, zebras. For the painted mule story to become a live issue the conversation would first have to be prepped accordingly. For instance, if the skeptic were to mention that a recent newspaper article had accused the zoo director precisely of such fraud, he could then convince us that, as we certainly wouldn't want to let ourselves be fooled in such manner, we'd have to admit we didn't actually know that we saw zebras. A second kind of contextualizing factor sneaked in here: At any time, we have economical—that is, practical—concerns that tell us, based on an analysis of possible costs and benefits, which error possibilities are worth investigating. If we deem it unnecessary to investigate some error possibility or other, we may fall back to a default position. Notice that in our current example the mentioning of an error possibility and relevant background information causes our stance toward one and the same position to change. At first we see no practical objections to making it a *default* position that what we see as zebras, *are* zebras. But then we learn about a news paper article that makes alternatives to this position (e.g. the mule hypothesis) suddenly sound just plausible enough to warrant investigation. The position that formerly enjoyed default status will from that point on either stay with us as a corroborated item of knowledge or suffer rejection as a falsehood. We could say that contingent events affect what it means to know *responsibly.* (Williams, 2001, p. 161) We might also say that what it means to know *for all practical purposes* depends on context.

There is a second kind of default position that Williams calls *methodological necessities.* These are positions that we must hold for a certain inquiry to be possible at all. Conversely, *methodological constraints* are kinds of doubt that cannot be pursued, lest the inquiry be unable to withstand the uncertainties they reveal. Together these make up the third kind of contextualizing factor. Williams draws on Wittgenstein here: Investigations require that, *qua* activity, we refrain from doubting some things (such as the existence of your measuring apparatus), not because we take them blindly on faith, but because it is

the condition of a researcher that he builds on prior assumptions. (Cf. Williams, 2004, p. 332) "What we are looking at is a function of what we are leaving alone," says Williams, so banning certain doubts "has nothing to do with either credulity or limited resources [but rather is] a matter of the *focus* or *direction* of inquiry." (Ibid.) Methodological necessities are untouchable until the subject changes his approach to solving the epistemic problem that occupies his mind. So, in just about any situation apart from epistemological discourse, if the skeptic asks how we know our measuring devices exist, we don't have to dignify his question with an answer. Here's how this helps Williams resolve issue (2): The existence of our measuring devices isn't postulated for *mere* convenience—it's not that we just don't feel like investigating their existence at this point—but is an ontological commitment that defines what we're doing. To doubt the existence of our apparatus is to change what we're doing—it is to change the subject. Paraphrasing an example that Williams borrowed from Austin may prove instructive:

Suzanne Arendt	:	Look over there. That's a goldfinch.
Jane Skeptic	:	How do you know?
Suzanne	:	By its red head.
Jane	:	But how do you even know it's a bird?
Suzanne	:	Because I can see it. (D'uh.)

Up to the third line, Suzanne and Jane seem to be engaged in ordinary ornithological discourse. But then Jane asks a bewildering question: How does Suzanne know she's looking at a bird? Eh, come again? Because she can see it, of course. (For a moment we wonder if Jane is asking such a question because something is blocking her view of the creature.) We might also answer that it's a bird because it has wings. Neither reply will satisfy the skeptic though: What Jane is trying to suggest is that we don't *know* that things are really how they appear to be. This concern introduces a radical shift in conversational topic: Jane, unlike Suzanne, isn't interested in the *characteristics* of goldfinches, but is instead looking for a general proof of the *existence* of goldfinches (or birds for that matter). In IC, what methodological necessities and constraints do, is to free the subject of the burden to prove that such boundary conditions hold—or at least they do until he takes up epistemology. (Williams, 2004, § 5)

We want to make special mention that Williams does not think methodological necessities are usually exhaustively cataloged. Williams explains that this is "why Wittgenstein insists that certain things are *in deed* not doubted" (2004, p. 332). Also noteworthy is that Williams believes methodological necessities and constraints determine the *disciplinary meta-context* "for all inquiries of a certain genre" (ibid.). We find these properties

interesting for they suggest that (1) a subject has certain commitments that he wholly shares with other subjects partaking in the same kind of research, and (2) it is decidable just-in-time what these commitments are.

The fourth kind of contextualizing factor hinges on intelligibility. Williams believes this point is closely related to the disciplinary meta-context, but is different in that it ranges over more contexts. Take the proposition "These are my hands." It would seem that this claim is incontestable in most contexts. A skeptic will need to steer the conversation carefully before his challenge will be recognized as meaningful. Perhaps he could trick one into saying "I know these are my hands because I can see them;" he could then point out that if we were brains in vats, we might still think we saw our hands, and therefore we don't know we have hands. Williams contends, though, that the skeptic would have only succeeded in making us take his challenge seriously by changing the subject to *knowledge as such*. But for Williams our belief in our hands, an external world, or in a historical past isn't a result of belief in our senses:

> These bedrock certainties derive their content—their meaning—from the particular practices of inquiry and justification that hold them in place. To believe in an historical, or an external world, just is to recognize certain types of error possibility, to demand certain kinds of evidence (in appropriate circumstances), and so on. Because such certainties are *semantically embedded* in our epistemic practices—thus unintelligible apart from them—it is a mistake to see those practices as justificationally dependent on the 'presuppositions' they embed. They are not assumptions because they are not, in the relevant sense, foundational at all. (Williams, 2001, pp. 165–166)

Crucial to understanding this position is knowing that Williams does not want to draw a sharp line between 'knowing that' and 'knowing how,' unlike how traditional epistemology would have it (ibid.). Williams points out that knowing requires skills such as recognizing relevant evidence, appropriate objections, and suitable replies, and that we learn these skills through our familiarity with paradigmatic examples. Therefore, the knowledge that we have hands is acquired at the same time we pick up the skill to know. When someone questions the proposition that we have hands, we don't think he or we may be mistaken, but we sooner suspect one of us does not know how to play the knowing game.

The fifth and last kind of Williams contextualizing factors stipulates that whether or not S knows that p depends on the factual situation. This type of factor reflects our commitment to an objective state of affairs when we claim knowledge. Sure, S might have all the reasons in the world to believe p at time t_1, but perhaps at t_2 a relevant, good,

new reason could show up for believing that (possibly) $\neg p$. IC has it that S is obliged to recognize such additional comments. In that case, the evidence cited at t_2 would force S to admit that he did not know p at t_1. And he *truly* wouldn't have known p, even if no one ever found out about this new evidence. Williams calls this the externalist element in IC.[3]

3.3. NON-SKEPTICAL PARADOXES

3.3.1. The Lottery Paradox

Next to providing remedies against skepticism, Brendel and Jäger identify solving the *lottery paradox* as an important goal of CC (2004, p. 148). The lottery paradox starts off with the observation that even though the chances at winning the national lottery are but slim, people will generally insist that, provided they bought a ticket, they do not *know* that they will miss out on the main prize. The paradox is that, oddly, lottery players will generally have no qualms about admitting to know they will never be multi-millionaires. What's going on here?

Cohen's explanation for this phenomenon is that at normal standards for knowledge, people *know* their lottery tickets won't make them multi-millionaires. By reflecting on the nature of lotteries, and the small chance that one has bought the winning ticket, however, the possibility of becoming a multi-millionaire becomes salient, and with this the standards for knowledge are raised. In fact, the new epistemic standards will become so high that, in this new context, the claim to know one will not hit the jackpot, is false. (Ibid.)

Williams has a different take on this paradox (2004, § 4). He thinks the trick to solving this problem is to look at it from a first person perspective. Someone playing the lottery should be expected to think he doesn't know he will lose because, quite simply, otherwise he wouldn't have bought a ticket. But on seeing a friend spend a lot of money on the lottery we may feel we know he is wasting his money. What Williams seems to say is that the lottery paradox confuses the story of two persons who evaluate the lottery game, each from their own context with different background information and practical interests. So it would seem that he is skeptical that someone can buy a lottery ticket thinking that he just might win and at the same time—though possibly in a different context—be sure that he'll always have to work for a living. Only when doing epistemology would we cook

---

[3]The summary of IC in this section is based on Williams, 2001, § 14; 2004, § 4ff, in addition to other cited sources.

up scenarios wherein both knowledge ascriptions are correct at the same time (in different contexts):

> [E]xamination of ordinary context-shifting intensifies our sense of the extraordinary character of the context supposedly created by "doing epistemology". Given the appropriate combination of background information and stakes, we can see how merely thinking of a defeater can raise epistemic standards. But this is a far cry from showing how standards can intelligibly be raised by reflection *alone.* (Williams, 2004, p. 330)

As such Williams's strategy with respect to the lottery paradox is to *dissolve* rather than *solve* it: He denies the lottery paradox is a true paradox.

Davis points out a second oddity concerning our knowledge claims about lottery games (2004, § 4). Suppose you asked two persons if you had won the national lottery. One of these persons, however, you show a news paper listing the name of the winner—and your name isn't mentioned. Both reply you didn't win anyway. Now, even though the odds of the news paper mentioning the wrong winner are much larger than the odds for any one particular person to win, Davis contends, we do intuitively feel that only the person who read the lottery results in the paper really knew you lost. A theory that can solve the lottery paradox discussed above, should also be able to explain this intuition. Williams offers the beginning of an explanation when he says this:

> The lottery case is [...] a case in which an error-possibility becomes salient *via* the interaction of background information and practical interests[.] When we enter a lottery, in the hope of winning a large sum of money, we decide then and there that nothing will count as conclusive evidence of losing short of the result's being officially announced. (Williams, 2004, pp. 329–330)

Unfortunately this reply doesn't address the problem that the odds that the newspaper would print the wrong name, are larger than the odds that the person who is listed as the winner would win (irrespective of his name being in the paper). Elsewhere, however, Williams points out that we delegate the justification of much that we believe to a larger social network (2001, § 4). Indeed: Why track evidence for and against evolution or global warming when you have scientists doing this for a living? We propose that in this second lottery problem, the trust that we place on the news paper should be seen as the outsourcing of the knowledge claim that, say, John Doe, not us, won this week's lottery. Of course this knowledge is fallible but we're confident that should the paper have made a mistake, a correction will be printed in its next issue. (And should we ever see our names printed as winners, we'd contact the lottery organization for confirmation.)

3.3.2. How a Sense of Urgency Can Affect Knowledge

Cohen observes that the standards by which we judge knowledge claims may vary as the importance of getting a correct assessment changes. Suppose someone, we'll call him Frank, is waiting for a train when someone asks him if it's true that the train to Brussels will arrive where they stand in ten minutes. As Frank is waiting for the same train, as he does every day, he spontaneously replies "Yes." But when the other person goes on to explain that it is very important that he arrives at his destination in time, Frank admits he doesn't know (for sure) if the train to Brussels is really about to arrive. He explains that he too just arrived at the platform and didn't consider yet that the train might be late or arrive on different tracks. (Cf. Davis, 2004, p. 228)

One explanation for Frank's change of mind is that he adopted higher standards for his knowledge claims when he found out much depended on them. Possibilities such as trains running late or being rerouted to different platforms became relevant at that point. (Cf. ibid.)

3.4. RECONSTRUCTIONS AND KNOWLEDGE

Let's discuss reconstructions that contain knowledge ascriptions. In what follows we require that reconstructions can contain propositions that state that some agent a knows that α. We will transcribe such propositions as "$\mathsf{K}_a\alpha$" but will not exhaustively define the semantics of the knowledge operator 'K.' Observe that on the classical account of knowledge, namely that knowledge is justified true belief, it should be a theorem that $\mathsf{K}_a\alpha \supset \alpha$. For instance, if $\mathsf{K}_a\alpha \in w_f$ (with w_f the world of facts) then $\alpha \in w_f$ because all known propositions are true propositions. Similarly, if an agent believes it to be knowledge that α then that agent must believe that α is true. Also, if a fictional story attributes to a that he knows α then the story represents α as true. And so on. Per a second effect of $\mathsf{K}_a\alpha$, α should also be true in another collection of worlds—namely those worlds that model belief systems as they are (re)presented in the relevant reconstruction containing $\mathsf{K}_a\alpha$. For instance, if $\mathsf{K}_a\alpha$ is true in the world of facts then the reconstruction(s) representing the actual beliefs of a will contain α. If $\mathsf{K}_a\alpha$ is true in the world standing for b's belief system then α will be true in the world representing what b thinks a's belief system is like.

'K' may seem too limited an operator because it does not allow us to express knowledge about reconstructions other than the world of facts. This, however, is symptomatic of what is at present a bigger shortcoming of our logic. Intuitively it is easily understood that

every proposition recorded in any reconstruction is a fact of some sort. For instance, if the proposition "Luke Skywalker has blond hair" is true in the reconstruction representing the Star Wars universe then the proposition "In the Star Wars universe, Luke Skywalker has blond hair" should be true in the world of facts. Presently though, we have no operator that can express propositions of the form "In w it is the case that α." If we had such an operator then we could encode knowledge that does not directly concern to the world of facts. Sadly, due to time constraints we cannot research how such an operator would precisely work. We are, however, confident that such an operator can be introduced in our logic and therefore presume that we can continue using our operator 'K' as described above.

When discussing epistemological contextualism one interesting kind of model of reconstructions has two worlds w_a and w_b that reconstruct a third world w_s. Usually w_s will be the 'real' world of facts. Also, reconstructions w_a and w_b should have been made by two agents a and b. We add that there is a relevant situation s that w_a and w_b truly reconstruct except for one or more K-statements (knowledge ascriptions) to which they assign different truth values. At the very least, the contextualist thesis implies that the different valuations for the knowledge ascriptions in w_a and w_b may all be correct at some point. That's about everything the different contextualist positions have in common though. Next we will look at conversational and subject contextualism. As we shall see, both families of theories complicate matters in their own distinct ways.

3.4.1. Conversational Contextualism

Conversational contextualists would say that different valuations for the same K-statements couldn't all be correct at the same time, even if the statements belong to different reconstructions. For consider a proposition $K_a\alpha$ in the setup just described. Specifically, if $K_a\alpha$ was evaluated in a conversation w_e (this would be a reconstruction too) then the proper truth value of $K_a\alpha$ would presumably be determined by the following factors:

(1) The conversation (w_e) in which $K_a\alpha$ is being evaluated. This world decides what standards apply;

(2) The evidence for and against α that a has or should reasonably have. We'll call this evidence situation s; and

(3) A reconstruction w_r containing all evidence for and against α that a has or should reasonably have. In the former case w_s could be used.

It would be sensible if w_r was reachable from w_e over R_{D} as the agents in conversation w_e intend to find out what a knows by reconstructing a's epistemic position w_r. This would be convenient because it would mean that in the scenario where this is the only world that w_e reconstructs, $w_e \vDash \mathsf{DD}\alpha \equiv \mathsf{D}\alpha$ will tell us whether a got the truth value of α right (or should have been able to do just that). Thus in such a simple scenario $w_e \vDash \mathsf{K}_a\alpha \implies w_e \vDash \mathsf{DD}\alpha \wedge \mathsf{D}\alpha$.

Having decided how we can tell from w_e if a got the truth value of α right, we will now discuss how we can deduce what standards of knowledge apply in w_e and how we can enforce these standards. Deducing what standards of knowledge apply is easy if we define these standards in terms of what error possibilities have been mentioned. We need a new operator to record what error possibilities have been mentioned. These error possibilities are stored in worlds which should diverge as little as possible from w_r except that they contain counterexamples to α. We define our new operator 'M' as follows:

$$\mathsf{M}\alpha \iff \text{there is a counterexample world in which } \alpha \text{ holds.}$$

In other words,

$$\mathsf{M}\neg\alpha \iff \text{a counterexample to } \alpha \text{ was mentioned.}$$

A new constraint $\mathsf{K}_a\alpha \implies \neg\mathsf{M}\neg\alpha$ can be introduced to enforce applicable standards.

This solution does not explain what makes for a counterexample but obviously it does not suffice to just say "There is a world where a is a zebra, where I see a, but where a does not exist." In fact, this is incoherent for one can only see things that exist. On the other hand, to say "You're not seeing a zebra" is a bare denial that is not informative with respect to how you could have been mistaken about seeing a zebra. What one should say instead is something akin to "There is a world where a is a zebra, where I *think* I see a, but don't *really* see a—in fact, a does not exist in this world and I'm really seeing b, a painted mule.[4]" So typically counterexamples will involve conceptual clarifications (e.g. that what one thinks one sees isn't always what one really sees) and thereby change what sort of statements can be used as premises in an epistemic proof (e.g. not "I see a zebra" but "I think I see a zebra and I'm an expert on zebras"). This subject however falls outside the scope of our logic of reconstructions.

---

[4]If painted mules are fake zebras then one indeed sees a 'fake zebra' when one mistakes a zebra for a painted mule. In the same sense a mirage can be seen, and may be said to exist, although the objects in it do not exist.

In a way w_r extends s, for whereas s lists a number of propositions, w_r adds truth values to these propositions and embeds them in a fully fledged world that can be used for logical reasoning. It's not clear, though, what is to be done with valuations in w_r for propositions that are not part of s. Some of these propositions may be incorrect or may, strictly speaking, not be knowledge. This is a problem because these propositions could be used for making inferences that are, from an objective point of view, invalid or otherwise not knowledge. Perhaps one solution would be to make worlds contain truth values only for a more limited set of propositions, such as those listed in a situation or the logical closure thereof. Whereas we have already discussed how valuations for a world can be limited with respect to objects, this solution would limit valuations with respect to predicate–object pairs. Alternatively perhaps '$\vDash$' could be extended to reason relative to a world *and* a situation. This approach would introduce situations as a proper part of our formal system—currently they are only useful as a device for reasoning about reconstructions on a meta level. A third option might be to no longer speak about reconstructions as worlds but about reconstructions as sets of worlds that are epistemic alternatives with regard to some situation. This option would take a clue from epistemic logic and would also introduce situations as first class citizens in our logic. Unfortunately time constraints prevent us from fully researching these tentative solutions.

3.4.2. Subject Contextualism

Subject contextualism does not let the truth value of knowledge ascriptions vary with the context in which the evaluation is made. Instead the context of the subject making the knowledge claim determines the criteria for knowledge. In other words, if w_r is a reconstruction that is intended to contain (be) knowledge then the context in which w_r was put together—we want to allow for some flexibility to what this means—fixes the context that partly determines how knowledge ascriptions are evaluated.

Suppose agent a makes a reconstruction w_a in which he claims that α. The question whether $\mathsf{K}_a\alpha$ holds is then determined by the context c in which a arrived at w_a. Now, c should contain all the evidence available to a for concluding α or $\neg\alpha$ in w_a—even if the evidence falls short of actually handing either conclusion. Thus c should be the set of propositions that make up the evidence for and against the knowledge ascriptions made by a in w_a. For a to know some proposition $\alpha \in w_a$ it is then required that a has a proof for α that requires only premises that are part of c.

In Barke's contextualism parties can, at any time, flag assumptions—here: premises—as suspicious. It is then up to the subject to either restore the assumption's credibility

or to give up his inquiry. To legitimate an assumption $\beta \in w_a$, of course, is to give a proof for β. Possibly c will contain everything required for inferring β. In this case the proof for β will have been mere epistemic information—its purpose will have been (1) to demonstrate to the challenger how β can readily be deduced from available evidence or even (2) to merely point out that some piece of evidence is actually available. However, the subject can only do this if the challenger does not bring with him new evidence against β. Alternatively the subject will need to start a new investigation (reconstruction) into β, thereby also bringing about a new context—which might also include the evidence against β that the challenger brought to the table (if any). Should the subject learn that it is indeed warranted to believe β then he can return to his original inquiry, reconstruction w_a, and context c. (Cf. Barke, 2004)

Williams's contextualism is more complicated. Let's go over the different kinds of contextualizing factors he identifies (see § 3.2). In the first place Williams embeds epistemic inquiries in conversations. That is, every reconstruction (that aims to be knowledge) is part of a conversation—just like in CC except IC has it there is at most one such conversation since it only cares about the conversation in which the reconstruction came about. Like before we want to define an M-operator although we now want to restrict it to *legitimate* criticism. Now even though Williams doesn't say every such criticism should be a counterexample as such, we do take it all such criticism—informally voiced—can be translated into a formal counterexample. That is, if in reply to someone's reconstruction of a zoo visit, we object that believing to have seen a zebra doesn't imply one was actually anywhere near a zebra, we will also be able to mention the counterexample where the zebra supposedly seen by this someone was but a figment of his imagination and the zoo had a painted mule instead. Thus we can leave the mechanics of our M-operator intact. One innovation we do require, however, is that whereas we found that in CC $\mathsf{K}_a\alpha$ entails $\neg\mathsf{M}\neg\alpha$, we now need $w_x \vDash \mathsf{K}w_a\alpha$ (for any world w_x) to entail $w_{a'} \vDash \neg\mathsf{M}\neg\alpha$—with $w_{a'}$ the privileged conversation that gave rise to w_a. In other words, we need operator 'K' to be able to reference reconstructions directly because we want to speak about agents possessing or not possessing knowledge at a certain point in time, in a given reconstruction. Also, AI systems employing this logic will need some procedure for deciding if some criticism is legit. This procedure will also need to take in account economical concerns about the costs and possible benefits of taking a piece of criticism seriously.

Like Barke, Williams acknowledges that investigations *qua* activities require that some assumptions are taken for granted. Williams calls these assumptions methodological necessities. Additionally Williams points out that many skeptical questions are unintelligible in most context (e.g. "Do I know these are my hands?"). Therefore, reconstructions, de-

pending on the kind of inquiry that induced them, have a whitelist of assumptions that cannot legitimately be challenged. AI systems will need to take this into account if they want to build on Williams's epistemology.

Finally, not only is knowledge fallible for Williams—we might be justified to say that $Kw_a\alpha$ even if $\neg\alpha$—but we can also not rule out that we don't have access to all information deciding if we're justified to say that $Kw_a\alpha$—so *justification* is fallible too. Consequently, there might always be more relevant M-statements than originally provided for in the conversation linked to a conversation. Philosophically we could say that some of these M-statements are simply invisible at first. However, when reasoning or when modeling conversations in real time, we might instead prefer to replace the conversation attributed to this or that reconstruction when a new M-proposition is discovered. This new conversation could be identical to the old one except that the truth value assigned to the M-statement should be flipped.

4. Conclusion

We believe one of the more interesting results of this master's thesis is that Kaplan's notion of objects of thought proves to need amendment. On the one hand this notion does capture the intuition that when we point to Jeff and say "You are wearing a business suit," and then point at Jeff's reflection and say "but he is not," we contradict ourselves—specifically, we describe a logically impossible situation. On the other hand, however, there also seems to exist another (stronger) notion of self-contradicting statements that is tied to rational thinking and acting. According to this second notion it would be irrational to say "Jeff is wearing a business suit and he is not wearing a business suit (at the same time)," but our earlier example using indexicals would not reveal an irrational thought if we were not aware that 'you' and 'he' referenced the same person. Thus our second notion of self-contradiction involves contradictions on a cognitive level—that is, it's more about invalid reasoning than about descriptions of logically impossible situations. Note that deciding if a contradiction is irrational involves tracking what references an agent assigns to dthat-terms. For instance, if we realized that we were talking about Jeff twice when we said "You but not he is wearing a business suit" then that would be irrational too. Consequently, not all forms of irrational self-contradiction can be detected at the level of cognitive significance of thought.

Another interesting result is that the genre of a reconstruction appears to be relevant in determining the referent of indexicals. This implies that 'now' does not always refer to the time at which a statement is made but can also indicate a—possibly fictional—time described in a story.

We have described our logic of reconstructions in much detail. Still, we have hinted that there are some gaps left that perhaps could be filled by future research. For instance, we have expressed the desire for an operator that allows us to evaluate propositions in arbitrary reconstructions of our choosing. Another topic that could be worth investigating is the relation between the belief operator and those reconstructions that represent what agents think. The most pressing problem, however, concerns the very foundations of our logic—viz. that it's not clear if reconstructions really are best formalized as worlds.

In much of our philosophical argumentation we have made use of situations to 'hide' those propositions about which a reconstruction had nothing definite to say. It's not clear, though, whether this approach is tenable in the long run. That's why we have suggested several alternatives—including the ideas of integrating situations into our logic, of replacing reconstructions by sets of worlds that stand for epistemic alternatives to said reconstructions, and of doing away with worlds altogether. Now, of course, in a sense this problem may indicate quite the fundamental flaw. However, at present we have no evidence that other results in this paper are at risk. Additionally, there's no reason to believe that foundational mistakes couldn't be fixed (should they occur).

Lastly, we found that epistemological contextualism provides devices for solving skeptical problems and other epistemological paradoxes. Our primarily research suggests that the different contextualist theories can be incorporated in our logic of reconstructions. We have, however, identified some technical innovations that such an enterprise would likely require. In any case, there remain sufficient details to be investigated to fill another fifty page report.

5. References

ALMOG, Joseph; PERRY, John; and WETTSTEIN, Howard (eds.). 1989. *Themes from Kaplan*. New York and Oxford: Oxford University Press.

BARKE, Antonia. 2004. "Epistemic Contextualism." *Erkenntnis*, 61, pp. 353–373.

BRAUN, David. Winter 2007. "Indexicals." In ZALTA, Edward N. (ed.) *The Stanford Encyclopedia of Philosophy*. URL http://plato.stanford.edu/archives/win2007/entries/indexicals/.

BRENDEL, Elke; and JÄGER, Christoph. 2004. "Contextualist Approaches to Epistemology: Problems and Prospects." *Erkenntnis*, 61, pp. 143–172.

CRESSWELL, M.J. 2004. *Adequacy Conditions for Counterpart Theory*, pp. 29–42. In Jackson and Priest (2004).

DAVIS, Wayne A. 2004. "Are Knowledge Claims Indexical?" *Erkenntnis*, 61, pp. 257–281.

DE VUYST, Jonas. 2006–2007. *Contextualisme in de hedendaagse epistemologie*. Bachelor's research paper, Vrije Universiteit Brussel.

FREGE, Gottlob. 1997*a*. *The Frege Reader*. Oxford: Blackwell Publishing.

FREGE, Gottlob. 1997*b*. *Thought*, pp. 325–345. In Frege (1997*a*).

GARSON, James. Spring 2008. "Modal Logic." In ZALTA, Edward N. (ed.) *The Stanford Encyclopedia of Philosophy*. URL http://plato.stanford.edu/archives/spr2008/entries/logic-modal/.

GERT, Joshua. 2007. "Vague terms, indexicals, and vague indexicals." *Philosophical Studies*.

HOFMANN, Frank. 2004. "Why Epistemic Contextualism Does not Provide an Adequate Account of Knowledge: Comments on Barke." *Erkenntnis*, 61, pp. 375–382.

HUGHES, G.E.; and CRESSWELL, M.J. 1996. *A New Introduction to Modal Logic*. London and New York: Routledge.

JACKSON, Frank; and PRIEST, Graham (eds.). 2004. *Lewsian Themes: The Philosophy of David K. Lewis*. Oxford: Clarendon Press.

KAPLAN, David. 1989*a*. *Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals*, pp. 481–563. In Almog et al. (1989).

KAPLAN, David. 1989*b*. *Afterthoughts*, pp. 565–614. In Almog et al. (1989).

QUINE, Willard Van Orman. 1980*a*. *From a Logical Point of View*. 3rd edition. Cambridge, Massachusetts, and London: Hardvard University Press.

QUINE, Willard Van Orman. 1980*b*. *Reference and modality*, pp. 139–159. In Quine (1980*a*).

ROMDENH-ROMLUC, K. 2006. "'I.'" *Philosophical Studies*, 128, pp. 257–283.

WILLIAMS, Michael. 2001. *Problems of Knowledge: A Critical Introduction to Epistemology*. New York: Oxford University Press.

WILLIAMS, Michael. 2004. "Knowledge, Reflection and Sceptical Hypotheses." *Erkenntnis*, 61, pp. 315–343.