Conceivability and the *A Priori*

Manfred.Kupffer@uni-konstanz.de

Dream the impossible dream
Joe Darion

0. Introduction

Does conceivability imply possibility? If yes, which kind of possibility? – These questions have enjoyed a lot of attention recently (see e.g. Yablo 1993, 2000, Chalmers 1996, 2002). David Chalmers' (1996) answer is that there is (i) a sense of "conceivable" in which conceivability directly implies metaphysical possibility and (ii) a sense in which conceivability implies what is sometimes called logical possibility and *epistemic* possibility at other times. (As I will use the term, something is epistemically possible if, and only if it is not knowable to be false *a priori*.)

Chalmers also famously thinks that often there is a road from epistemic possibility to metaphysical possibility, witness his version of Kirk's (1974) Zombie Argument. In the following we will not be concerned with the question whether the Zombie Argument is sound, though. However, we will deal with the question how conceivability is employed there.

I will show that there is reason to doubt Chalmers' ambiguity thesis. At least there has to be a third sense of "it is conceivable". Even in the context of what seems to be the most natural understanding of the Zombie Argument, none of his two senses are appropriate. This serves to discredit the ambiguity thesis and, more generally, the idea that conceivability always implies *some* kind of possibility. In fact I will try to show that "it is conceivable that S" merely implies that it is epistemically possible that S is not inconsistent.

A caveat: conceivability, as ordinarily understood, does not even imply that much. Maybe it has no logic worth speaking of. E.g., I can conceive of trisecting the angle with compasses and ruler alone although it is analytically true that there is no such construction. Note, that such a construction would be no longer conceivable when under the scrutiny of ideal rational reflection. Therefore it pays off if we turn from conceivability in the ordinary sense to an idealised notion. Following Chalmers we will only deal with *ideal rational* conceivability, where something is ideally rationally conceivable if, and only if, it is *prima facie* conceivable and sustainable under ideal rational reflection. The move should be a familiar one to anyone interested in the logic of the attitudes. Is there anything worth being called the logic of belief? – Only if we are dealing with a severely idealised and rationalised notion of belief.

A second *caveat*: because conceivability can be defined as the possibility of conceiving, "conceivable" shares almost all of the interpretive vagueness of "possible". Of course context will sometimes remove that vagueness a bit. Possible ways to make the "able" in "conceivable" precise are all broadly metaphysical;

sometimes restricted versions of metaphysical possibility are best suited, e.g. feasibility for us humans, or even feasibility for *me*. In the following it is mostly safe to assume that it is metaphysical possibility in the widest sense that is intended. This also seems to be the sense Chalmers intends. But keep in mind people often mean more than that when they use the word.

I will first inspect the two senses Chalmers proposes, taken on their own. The third section, then, deals with his ambiguity thesis, the fourth with the sense of conceivability appropriate for the Zombie Argument. While the first part of the paper deals with the logic of conceivability, the second part is concerned with its semantics. We will be looking for an analysis of "conceivable" that does not imply any kind of possibility. Surprisingly we end up with a variant of Chalmers' own analysis; the only important disagreement seems to concern not the analysis as such, but its proper interpretation. In the sixth section I will show how to square the weakness of the resulting notion with the fact that thought experiments sometimes inform us of what we should rationally believe to be possible. Finally, at the end you may find an appendix that presents the main definitions and results wrapped in the familiar language of diamonds and boxes.

1. Blind Optimism

I will use the name "Blind Optimism" to refer to the view that conceivability implies possibility in its philosophically most familiar sense, i.e. metaphysical possibility. Blind Optimism is false if there are necessary falsities that are conceivable[1]. Take (1) below.

(1) Hesperus ≠ Phosphorus.

It is widely held that (1) is necessarily false and I will simply accept this view here. (For an argument see Kripke 1972). In Kripke's story these names are introduced independently, and prior to empirical investigation it is not clear that they co-refer. So at that early stage (1) is conceivable. But then, sorry, Blind Optimism is mistaken.

2. Vulgar Kripkeanism

Note that although (1) is metaphysically impossible, it is epistemically possible. Consequently, Vulgar Kripkeanism holds that while conceivability may not imply metaphysical possibility, it still implies epistemic possibility. But is this true? – It could be refuted by an example of something that is conceivable although it is a denial of an *a priori* truth. Consider

(2)   Everything is as it actually is.

---

[1] In the following I will frequently say things like "S is conceivable", "S is possible", or "S is *a priori*" in order to abbreviate "It is conceivable that S", "it is possible that S", and "it is knowable *a priori* that S", resp.

I will argue that (3) and (4) are true.

(3)   It is knowable *a priori* that everything is as it actually is.
(4)   It is conceivable that it is not the case that everything is as it actually is.

If I am right, Vulgar Kripkeanism is mistaken.

The case for (4) is easily made. Of course it is conceivable that not everything is as it actually is! Why, it is even conceivable for me that my desk is tidy! It is so in the sense of *prima facie* conceivability, of course. But *prima facie* conceivability only fails to make it for ideal rational conceivability if an ideal agent is able to detect it contains some hidden contradiction. Now, in the present example there is no hidden contradiction to be detected.

You might object that the negation of (2) is still ideally rationally *in*conceivable because it contains itself a (not so hidden) contradiction. It is inconsistent because, necessarily, if you say it, it is false. I contend that "one cannot assert it truly" may be a legitimate and useful sense of "inconsistent". But it is not the only sense. And it is not even the relevant sense here, where the question is whether some prima facie conceivability is logically sound. When conceiving we simply don't *assert* things! — So what *is* the relevant sense of inconsistency? I think analytical falsity is the obvious choice. Now here is something I am going to assume throughout the paper: *the relevant sense of logical truth is analyticity*. Hence, inconsistency is understood as analytical falsity and the negation of (2) is not inconsistent in that sense. And this is why I can maintain it is ideally rationally conceivable.

While the case for (4) may be easily made, (3) is in need of additional argument. The problem is that an opponent could try to resist (3) simply for the reason that (4) is true and conceivability implies epistemic possibility. If we want to win the argument and maintain that (2) is *a priori* we need a substantial independent explication of the *a priori*. Unfortunately, if we provide one, we lay ourselves open to attack: there are many possible explications of this rather vague notion, and the opponent may well reject the one we chose. I will try to circumvent this calamity by choosing a notion that is fairly general and has rather inevitable features. So which notion do I have in mind?

Of course a notion of ideal, rational knowledge *a priori*. This is so, since we want to link it to ideal, rational conceivability. But there are several such notions available.

I will not choose Kaplan's (1989) explication of the *a priori* in terms of two-dimensional semantics, because, as a redefinition of an epistemological notion in purely semantic terms, it is itself in need of justification. It should be noted, though, that together with the standard semantics of "actually", Kaplan's definition is in accordance with (3).[2]

I will not choose to explicate knowledge *a priori* as knowledge someone possesses solely in virtue of being a knower (and hence, regardless of what the knower has experienced). Maybe (3) is true in this sense of *a priori*; maybe it's not. This is hard to

---

[2] Kaplan treats "*a priori*" as a predicate of object-language sentences, though. But surely a sentence S is *a priori* if, and only if, "it is knowable *a priori* that S" is true? What else could "S is *a priori* mean"? – It is easy to see why Kaplan cannot like this reasoning, though, for it would mean that "it is knowable *a priori* that" is a *monster* (see Kaplan 1989) and Kaplan wants to avoid monsters.

decide. Take your dog. She's a possible knower, cause she actually knows things. She knows where she dug the bone. But does she know that (2)? This is hard to decide, because there is no non-verbal behaviour we could take as evidence for or against her knowing that (2).

I will also not choose to explicate knowledge *a priori* as knowledge in virtue of being able to experience. For one thing, this does not remove the problem about (2) and your dog. And, *pace* Kant, it is not clear that very much can be got out of this notion. Instead I will explicate knowledge *a priori* as knowledge had solely *in virtue of semantic competence*. In this sense A *knows a priori* that S, if and only if, A knows that S simply in virtue of knowing the meanings of the expressions A knows. A sentence S of English may be said to be (knowable) *a priori* if and only if it is possible that someone knows *a priori* that S. It is easy to see that (2) is *a priori* in this sense. Everybody who knows the meaning of (2) knows that (2) is true. But if a person knows that (2) is true and knows the meaning of (2), then we are entitled to say that the person knows that (2). It follows from the above definitions that the person knows that (2) *a priori*. It also follows that (2) is *a priori*.[3]

If you want to investigate the logic of the attitudes make sure not to mix different readings. It would be very easy, and indeed, too easy, to find counterexamples to Vulgar Kripkeanism if we were allowed to change readings during the race. E.g. it is *a priori* that the man with the brown hat$_{de\ dicto}$ is the man with the brown hat$_{de\ dicto}$, although it is conceivable that the man with the brown hat$_{de\ re}$ (= Ortcutt) is not the man with the brown hat$_{de\ dicto}$. – Nobody should accept this example as a counterexample to Vulgar Kripkeanism. So I should finally show that I did not change readings during the race.

I will show that, in my counterexample, both embedded occurrences of (2) may be understood in a uniform, *de dicto* way; effecting a narrow scope reading of "actually", in which the ascriber does not use "actually" to refer to the world the ascriptions are made in (instead "actually" is somehow shifted to worlds "considered as actual" by the subject of the attitude).[4]

Let us first treat (3). On a *de re* reading where "actually" is taken to refer to the actual world, (3) claims that the precise nature of the actual world can be known *a priori*. (3) is false, according to that reading. But there is another reading, a reading in which (3) is true, and indeed a mere triviality. This is the reading I have used in my above argument for (3). There, assent to (2) implies that the subject knows that (2), regardless of what the precise nature of the actual world turns out to be, so the reading is a *de dicto* reading.

Now I finally need to show that there is an analogous reading of (4), and that (4) is true according to it. Consider what I said in favour of (3). I can conceive that not everything is as it actually is. I can do it by imagining a situation where my desk is

---

[3] In Kupffer (2003) I defend the present account of the *a priori* in greater detail. I also show how to extend the treatment of sentence (2) to sentences containing perspectival indexicals like "I", "here" and "now". It turns out that e.g. "I exist" is also *a priori*. (Yablo 1993 argues that conceivability cannot be equated with epistemic possibility because "I do not exist" is conceivable while "I exist" is *a priori*. But he fails to argue for the latter claim.*)*

[4] There is also a less important scopal issue about "everything", which should be interpreted to have narrow scope, too.

tidy. This scenario is telling because *I think that the desk is in a mess*. Now, according to the reading I have in mind, this is all I have to do in order to conceive of the opposite of (2). It is not necessary that my desk *really* be in a mess. Suppose, unbeknownst to me, someone has tidied it. Still there is a sense in which I have managed to conceive that not everything is as it actually is. – This sense is a *de dicto* reading of (4), too, a reading according to which "actually" does not refer to the actual world. Hence, in my counterexample I have managed to stick to the same reading in (3) and (4). It is therefore, that the counterexample has bite.

3. Chalmers' Thesis

Now let's turn to what may somewhat maliciously be termed a combination of Blind Optimism and Vulgar Kripkeanism. According to Chalmers, there is a kind of conceivability that implies epistemic possibility and another kind that directly implies metaphysical possibility. Using his own terminology, we may put his proposal in the following form.

<u>Chalmers' Thesis</u>: n-conceivability only implies n-possibility (n $\in\{1,2\}$)[5]

1-possibity is what I have called epistemic possibility above. 2-possibility is metaphysical possibility. Given these two kinds of conceivability, conceivable necessary falsities like (1) may be dealt with by appeal to 1-conceivability, whereas conceivable epistemic impossibilities may be dealt with by appeal to 2-conceivability. This means Chalmers really needs at least those two senses.[6] But are they enough? Take, e.g.
    (5)   Neither is everything as it actually is nor is Hesperus Phosphorus.

(5) is the conjunction of an epistemic impossibility and a metaphysical impossibility. Hence (5) is both 1-impossible *and* 2-impossible. But, at an early stage prior to empirical investigation, it used to be conceivable that Hesperus is not Phosphorus; it also used to be conceivable then that not everything is as it actually is. Suppose we are living at that early stage. Probably the specific scenarios one imagines when conceiving of these two things are independent and hence combinable. Therefore, also (5) is conceivable. Since (5) is not analytically false, it is also not inconsistent in the relevant sense. Hence, it is even ideally rationally conceivable. But then, here is a case of conceivability that cannot be accounted for by either of Chalmers' two senses. We are at least in need of a third sense of "conceivable".

---

[5] The thesis is most clearly held in Chalmers (1996, p.67); Chalmers (2002) argues only for n=1, the n=2 part of the thesis is not further discussed and mainly termed "plausible".
[6] Chalmers (1996) argues one gets this ambiguity for free, because it is due to a systematic structural ambiguity; modal operators might apply either to primary or secondary propositions.

4. The Zombie Argument

And there is another case that calls for such a third sense, namely the Zombie Argument. Only that this time the story is somewhat more complicated. The Zombie Argument is an argument from conceivability to (metaphysical) 2-possibility.[7]

> The Zombie Argument
> It is conceivable that
> (i)   physically, everything is as it actually is but
> (ii)   there is no phenomenal consciousness
> _____(by the use of a two-dimensional apparatus:)[8]
>
> It is 2-possible that
> (i) physically, everything is as it actually is and either
> (ii) there is no phenomenal consciousness or
> (iii) some other positive fact of the actual world is absent.
> (In short, *materialism is false*.)

I will not deal with the argument as such, here. Instead I will confine myself to the interpretation of the premise. How is it to be understood?

Let me first deal with the readings connected to the presence of "actually". The premise can again be understood in a *de re* and a *de dicto* way. Understood *de re*, the premise means that it is possible to conceive of the actual state of the world in all its physical detail together with the absence of consciousness. Whether this is true depends on what is meant by "possible" here. If it is something like "feasible for humans", the premise is false since it is not feasible for any human to conceive of all the physical details of the actual world (not even of all the physical details of a single conscious individual at a time). But if metaphysical possibility is at stake, then, maybe, the premise is true, even understood *de re*. The problem is that we can't tell! At least we cannot establish the possibility of conceiving that (i) & (ii) by trying to imagine it ourselves, because, as I have said, this is beyond our range. So, if the premise is to be understood *de re*, the premise itself lacks intuitive appeal. It would rather be something that has to be argued for. But the premise has intuitive appeal and has caught the imagination of many! When the argument catches our imagination, we take it to ask us to imagine a certain thing, namely that (i)&(ii). How do we react? "Sorry, I can't; I don't know the world in all its physical detail"? Certainly not. Instead we're trying to imagine what we are told, while it doesn't matter to us at all how little we know of the actual physical goings-on in the world. And at least some say they actually can imagine such a thing. For them, their imagination proves the premise is true in the ordinary sense of conceivability. And this may be a reason to think the premise is also true in the ideal rational sense Chalmers aims at. Anyway, how we react shows that the sense in which the premise of the Zombie Argument has intuitive

---

[7] Chalmers (1996), pp. 94, 123 and 131-134. My reconstruction is, as far as it goes, identical to the one in Brueckner (2001), p.187.

[8] The relevant details of Chalmers use of two-dimensional modal logic are presented informally e.g. in Brueckner (2001) under the name of "Chalmers' answer to the Standard Objection".

appeal cannot be a *de re* reading in which the actual physical goings-on enter into the truth conditions of the premise. Instead it is understood *de dicto*, such that "actually" is not taken to refer to the actual world, but rather to worlds taken as actual.[9]

Now let us turn to the notion of conceivability at work when we understand the premise that way. Is it 1-conceivability? – It can't be. When we are imagining a situation in which physically, everything is as it actually is *but* there is no phenomenal consciousness, we are imagining a situation in which *it is not the case* that everything is as it actually is, i.e. we are imagining a situation in which something epistemically impossible is true. But then, the imaginability of a situation of that kind can't be taken to imply epistemic possibility. Therefore, 1-conceivability is ruled out.

2-conceivability is ruled out, too; all hands agree that the argument tries to establish dualism purely by *a priori* methods. But then, even the Zombie Argument (on its most natural reading) calls for a notion of conceivability that differs from Chalmers' two notions.[10]

5. On the analysis of "conceivable"

In this section I will try to find a notion of conceivability that is appropriate for cases like (5) and the Zombie Argument. It will turn out we find an all-purpose notion.

Let us start with *conceptions*. No conceiving without a conception. Conceptions are mental stories; sometimes, in the case of perceptual imagination they are mental images. In the following I will confine myself to non-perceptual imagination, though. If we conceive that S we form some conception that we believe to be coherent and to imply S. This motivates the following chain of definitions. Let "A has S-conception C" mean: C is a conception, A forms C, and A believes C to be consistent and to imply S. Then we can say that A *conceives* that S if and only if A has some S-conception, and that S is *conceivable* if and only if there is a possible A that conceives that S. All this can be understood in two ways, depending on whether we employ a notion of *prima facie* belief, or the kind of belief that will survive ideal rational reflection. (The first one is what we ordinarily call "belief".) I will call the resulting notions *prima facie* conceivability and *ideal rational* conceivability, accordingly.

This is already all I will need in terms of an analysis of conceivability. But before I apply it, let me briefly compare it with what may be termed the "standard analysis"

---

[9] To say the *de dicto* reading is the most natural one does not mean, however, it is the reading Chalmers himself had in mind. At least nowadays he prefers a non-indexical version of the premise (see Chalmers 2002, p.196) that is equivalent to the *de re* reading discussed below.

[10] An additional reason to doubt that 1-conceivability is appropriate for the Zombie argument could be the following. In Chalmers (1996) it is claimed that 1-conceivability is an operator on primary propositions. Then it would be closed under conjunction with *a priori* truths, i.e. if T is *a priori*, then it is 1-conceivable that S&T if and only if it is 1-conceivable that S. This is so because, in such a case the primary propositions of S and S&T coincide. So, analysed in terms of 1-conceivability, (i) above would be redundant and the premise of the Zombie-Argument would be equivalent to

(6)   It is conceivable that there is no phenomenal consciousness.

Chalmers certainly did not want to argue from that simple premise alone! – Perhaps it would help with this problem to switch to a situation semantic treatment of conceivability (this would be consistent with Chalmers (2002)).

of conceivability (Yablo 1993, Chalmers 2002). It will turn out that the two analyses are equivalent.

The standard analysis starts with the act of *conceiving*. According to that analysis, to conceive that S is to *imagine a situation* that one takes to verify S. If this is correct, imagining has an objectual character i.e. it is a relation to some situation. So far the differences between the two accounts are striking; whereas my analysis is in terms of conceptions, the standard analysis is in terms of situations. But Chalmers admits the objectual character is *mediated*, the relation to the situation is mediated by some means of representation. E.g. in the case of perceptual imagination one forms some mental image that represents the situation one imagines. There is also a non-perceptual kind of imagination, which Chalmers calls *modal imagination*. Modal imagination is the kind of imagination that is relevant for our purposes. Unfortunately Chalmers is less explicit about this more important kind. E.g. he does not mention in what the intermediate representational means consists in. But I think we can safely say that it consists in conceptions. So to modally imagine a situation is to entertain some conception; in fact my analysis is the result of a simple translation of the standard analysis into the language of conceptions based on this insight. Compare my analysis above with the following statement of the standard analysis.

> S is *prima facie* positively conceivable when one can modally imagine a situation that one takes to be coherent and that one takes to verify S. S is ideally positively conceivable when S is *prima facie* positively conceivable and this positive conceivability cannot be undermined on idealized reflection. […] S is ideally positively conceivable when one could coherently imagine a situation that verifies S. Chalmers (2002), p.153

But while my analysis may simply be a way to express the very same thing, I still think that here, the language of conceptions is superior to the language of situations. First it is not clear how coherency, in the logical sense it is understood here, can be predicated of situations at all. Situations are "configuration[s] of objects and properties" (p.151). So how can one "take a situation to be coherent"? Coherence (consistency) is a property of the conception, not of the situation conceived! Also, if conceptions represent situations, what the subject *takes* the situation to verify depends on the conception, not on the situation. My analysis makes this dependence explicit. Third, the question arises what is meant by "verification". Of course it could be something like the truth-in-a-situation-relation familiar from situation semantics. But Chalmers says instead it is analogous to "*a priori* entailment" (p.152). Now entailment is a relation between (sets of) sentences, not a relation between a situation and a sentence. So even here, it would be wise to replace situations by conceptions. – Nothing I have said so far precludes that it may be a characteristic feature of conceptions that they somehow represent situations. (Neither does it rule out that the correct logic for my account could be situation semantic rather than classical.)

So what does conceivability imply? –This can be easily seen. Since it holds that if C is consistent and C implies S, then S is consistent, it also holds that if A believes C to be consistent and to imply S, then A believes that S is consistent, at least if belief is treated in an ideal rational sense. Therefore, *if A conceives that S, then A believes that S is consistent*. Now I have a minor addition to make to the analysis developed so far. If A

conceives that S the analysis so far does not demand that A bears the attitude to S[11] in its ordinary sense. I think this is a mistake. If, e.g., Fritz thinks that "Goldbach's conjecture is true" means "pigs can fly" we cannot conclude that Fritz can conceive that Goldbach's conjecture is true from his forming some specific conception of flying pigs and taking it to imply the sentence "Goldbach's conjecture is true". So A conceives that S if, and only if A *knows the meaning of S* and has some S-conception. Therefore, if S is ideally rationally conceivable, then there is some possible competent speaker A that conceives that S; and we have seen above that this implies that A believes that S is consistent. But to say that there is a possible competent speaker that believes that S is consistent means that it is epistemically possible that S is consistent. Hence, *if S is ideally rationally conceivable, then it is epistemically possible that S is consistent.*

Perhaps this result could still be strengthened.

I you want to go beyond the mere epistemic possibility that S is consistent, you might employ the following auxiliary postulates.
(A)   If it is epistemically possible that S is consistent, then S is consistent
(B)   If it is epistemically possible that S is consistent,
            then it is epistemically possible that S is metaphysically possible.
If (A) is true, then conceivability implies consistency, if (B) is true, it implies the epistemic possibility of metaphysical possibility. I will not try to decide whether (A) and (B) are true, here. Let me merely list pros and cons.
Contra (A): Analytical falsity is our preferred candidate for the explication of the notion of inconsistency. Now (1) is often said to be analytically false. Since it is also epistemically possible that (1) is consistent, the proposed analysis of inconsistency predicts that (A) is false.
Pro (A): It is not really clear that (1) is analytically false. E.g. the latter claim is not borne out by causal descriptivist semantics.
Contra (B): One could construe formal models that falsify (B). Note that an individual can well believe that a sentence S is consistent but impossible, e.g. if S is an *a posteriori* necessity like (1). Therefore we could construe a model in which *every competent speaker* believes that some sentence CI is consistent but impossible. In such a model it is true that it is epistemically possible that CI is consistent but *a priori* that it is impossible that CI, i.e. (B) is false.
Pro (B): The counterexample looks at best artificial. If (B) is false, then *no* competent speaker in his own mind could think that it is possible that CI. But the only examples of sentences no competent speaker could reasonably believe to be possible that I can think of are sentences that are known to be analytically false by every competent speaker.

Anyway, it is important to notice that our result *cannot be strengthened to the extent that conceivability implies any kind of possibility.* This is so because the epistemic possibility of consistency does not imply that either. There are clear cases of sentences which are not possible in any of our two senses but which competent speakers believe to be consistent. (5), repeated here for convenience, is such a case.

(5)   Neither is everything as it actually is nor is Hesperus Phosphorus.

---

[11] Or to a translation of S. – In the following I am going to ignore the possibility that what is ascribed is merely a translation of the sentence the subject bears the attitude to.

Even for competent speakers, (5) used to be believably consistent (not analytically false) at an early stage of the introduction of the two names; but (5) was neither epistemically nor metaphysically possible, then.

How come Chalmers, who has basically the same analysis of conceivability, still thinks conceivability implies possibility? I have a tentative diagnosis to offer. At least I am able to explain his predictions about 1-conceivability. Cast in the language of conceptions, our disagreement seems to be about the interpretation of the expressions "A believes C to be consistent (coherent)" and "A believes C to imply S". It is rather obvious that one should understand them in terms of what they are composed of, namely the notion of belief and a logical notion. But how are these logical notions in turn to be understood? What Chalmers proposes in effect comes down (i) to equate logical truth with apriority (and, consequently, consistency with epistemic possibility and implication with the apriority of the conditional) and (ii) to treat *a priori* as *exportable* (i.e. if an ideal agent believes that S is (not) *a priori*, then it is (not) *a priori)*. If we combine (i) and (ii) with our analysis, this has the effect that conceivability implies epistemic possibility. – I think at least (i) is a mistake. Briefly, contingent truths *a priori* are not logically true, their denials not inconsistent. E.g. neither the negation of (2) nor "I do not exist" nor "I am not here" are inconsistent. Why, they are possibly true!

At least for the purposes of our analysis, logical truth should not be equated with apriority. Above I have opted for analyticity. Accordingly, "C implies S" should be taken to mean that C->S is analytic, and "C is consistent" that ¬C is not analytic. No sentence is analytic if its negation is possibly true; hence (5), the negation of (2),"I don't exist", and "I am not here" are consistent on the proposed explication of consistency. However, for the purposes of our analysis it really doesn't matter how consistency is interpreted precisely, as long as it does not turn out to be both exportable and to imply some kind of possibility.

Now, since our analysis of conceivability does not imply any kind of possibility, we have already got what we need: a notion of ideal rational conceivability that is appropriate for all the examples discussed so far. Therefore, 1- and 2-conceivability are no longer needed as separate senses.[12] We can do with the one weak sense we've got alone. Indeed we should; in the absence of firm intuitions of ambiguity a non-ambiguity analysis is almost always preferable.[13] Summing up: *there is only one kind of conceivability, and it does not imply any kind of possibility.*

---

[12] Of course, if we want to have something like 1-conceivability for theoretical reasons, we can *define* it from our notion. Let us say S is 1-conceivable if "actually, S" is conceivable. Interestingly, something like this can be found in Chalmers' informal characterisation of 1-conceivability. "We can say that S is *primarily conceivable* (or *epistemically conceivable*) when it is conceivable that S is *actually* the case". Chalmers 2002, p.157. This quotation strongly suggests that 1-conceivability should be defined in terms of a more basic notion of conceivability. Unfortunately, Chalmers fails to provide one.

[13] The literature also knows a distinction between positive conceivability (requiring a conception) and negative conceivability (no conception required, what is conceived is merely not ruled out). I tend to think this is not a case of a genuine ambiguity but of different ways context makes the "able" in "conceivable" precise.

6. Thought experiments

So far it was the weakness of our notion of conceivability that seemed to be its biggest advantage. It made the notion applicable both to necessary falsities *a posteriori* and to contingent *a priori* falsities. But perhaps this very weakness gives reason to doubt the analysis. The problem is that, sometimes, we seem to gain knowledge about what is metaphysically possible while sitting in the armchair. At least thought experiments often lead us to believe things about what is metaphysically possible. If my account says they don't, something has gone wrong.[14]

As an example let us take a look at TWE the Twin Earth experiment (see Putnam 1975). TWE tries to establish that (7) below is possible. If it is, we may conclude that reference does not supervene on what's in the head.

(7)   'Water' refers to something different from what it actually refers to while Oscar's narrow psychological state remains what it actually is.
(The narrow psychological state of Oscar is the psychological state of Oscar *taken in isolation*.)

Like the conceivability arguments above, TWE rests on a particular story C. This time it is the (modal version of the) tale of twin earth. It is so widely known that it need not be told here. Let's accept an abbreviated form and take our C to be the following two-part conception.

(i) in Twin World, a world that is identical to the actual world wrt. the narrow psychological state of Oscar, something different from H_2O plays the water role, although (ii) actually it is H_2O that does.

What happens if we perform TWE, according to my account? We form C and take it to be coherent and to imply (7) (presupposing that in a world w "water" refers to whatever plays the water-role in w). Given what I have said so far this only implies the belief that (7) is *consistent*. – But this is clearly not enough. TWE is supposed to inform us of our commitment to believe that (7) is *possible* (and it does, given we buy into an appropriate story about how reference is fixed as well as into the notion of a narrow psychological state). So is my account too weak?

Everything is all right. My account is correct, and this can be squared with the fact that thought experiments induce (or unravel) beliefs about what is metaphysically possible. The solution of the problem is, in a nutshell, that our C has special properties that induce the stronger belief. C is such that if an ordinary analytic philosopher believes that "H_2O plays the water role" is true and takes C to be consistent, she is also ideally rationally committed to believe that "Possibly, C" is true. The following is an informal argument for this claim.

Let A be an ordinary analytical philosopher, and C be A's (7)-conception. Let us also assume that A is a competent speaker (of all the expressions to follow). Observe it is not *a priori* that "Possibly, C" is *false*. (Above I have speculated whether this holds for any sentence that is believed to be consistent.

---

[14] This criticism has been raised in Yablo (1993) against various weak notions of conceivability.

However this may be, it certainly holds for C.) Indeed the ordinary analytic philosopher A has no reason to believe that "Possibly, C" is false. (She believes that H_2O plays the water role, she does not believe that "plays the water role" is a rigid designator, etc.) Therefore it is compatible with her beliefs that "Possibly, C" is true, i.e. there is a way the world might actually be as far as A's beliefs go according to which "Possibly, C" is true. But third, because of A's competence, whether it is true according to some such way should not depend on any special features of that way, other than that "H_2O plays the water role" is true according to it. Now the latter is something that is true according to *all* of the ways the world might actually be as far as A's beliefs go. That implies that if "Possibly C" is true according to one such way, then it is true according to *all* such ways. Hence "Possibly C" is true according to all such ways, i.e. A believes that "Possibly, C". That's what we wanted to show.

So our ordinary analytical philosopher A believes that "Possibly, C" is true. Since A also believes C to imply (7), A additionally believes that "Possibly, 'water' refers to something different from what it actually refers to while Oscar's narrow psychological state remains what it actually is" is true. Because she is a competent speaker, it finally follows that she believes that it is possible that (7). And that is why TWE can be used to persuade people that (7) *is* possible and that therefore, reference does not supervene on what's in the head.

The same kind of argument can't be run for cases like (1), for instance. It is for this reason that forming a (1)-conception does not lead to the belief that (1) is possible.

Let's again assume we are at the stage prior to the empirical discovery that Hesperus is Phosphorus. Let D be your (1)-conception and let us assume it is *compatible* with your beliefs that "Possibly, D" is true. Even then, one cannot already infer you *believe* "Possibly, D" is true. That inference is blocked, because whether "Possibly, D" is true according to a particular way the world might actually be as far as your beliefs go, *depends on the features of that way*. The reason is the following. "Possibly, D" implies "Possibly, (1)". But, as a believer in rigid designation, whether "Possibly, (1)" is true according to a way the world might actually be as far as your beliefs go, depends on the reference of "Hesperus" and "Phosphorus" according to that way.

We have seen that it depends on the conception, our linguistic competence, and on our other beliefs whether a thought experiment really leads to believed possibility. Sometimes it does; then it informs us of our commitments to think something possible. And that's why thought experiments are of great philosophical importance.

Appendix: In terms of diamonds and boxes

<u>Primitives</u>
$\text{Comp}_x S$: x is a competent speaker of S
$\text{Bel}_x S$: x believes that S
$\text{Has}_x S$: x has conception S[15]
L   : logical truth
         : metaphysical necessity

<u>Definitions</u>

Possibility	$\Diamond S$	$:\Leftrightarrow \neg\ \neg S$
Consistency	CS	$:\Leftrightarrow \neg L \neg S$
Implication	$S' \models S$	$:\Leftrightarrow L(S' \rightarrow S)$
Apriority	$_e S$	$:\Leftrightarrow \forall x\,(\text{Comp}_x S \rightarrow \text{Bel}_x S)$
Epistemic possibility	$\Diamond_e S$	$:\Leftrightarrow \Diamond \exists x\,(\text{Comp}_x S \wedge \text{Bel}_x S)$
Conceives	$\text{Conc}_x S$	$:\Leftrightarrow \text{Comp}_x S \wedge \exists S'(\text{Has}_x S' \wedge \text{Bel}_x (S' \models S) \wedge \text{Bel}_x CS')$
Conceivability	$\text{Conc}\ S$	$:\Leftrightarrow \Diamond \exists x\,\text{Conc}_x S$

<u>Postulate</u>[16] $(\text{Bel}_x (S' \models S) \wedge \text{Bel}_x CS') \rightarrow \text{Bel}_x CS$

<u>Theorem</u> $\text{Conc}\ S \rightarrow \Diamond_e\ \text{Bel}_x\ M\ S$
               (follows directly from the definitions and the postulate).

<u>Kaplanian semantics</u>[17]
For every sentence S, $|S|$, the character of S, is a function from pairs of contexts c and worlds w into the set $\{0,1\}$.

Necessity	$|\ S|(c,w) = 1$	$:\Leftrightarrow$ for all w: $|S|(c,w) = 1$,
Logical truth	$|LS|(c,w) = 1$	$:\Leftrightarrow$ for all c, w: $|S|(c,w) = 1$,[18]
(1)	$|(1)|(c,w) = 1$	$:\Leftrightarrow$ Hesperus $\neq$ Phosphorus,
(2)	$|(2)|(c,w) = 1$	$:\Leftrightarrow$ in w, everything is as in the world of c.

In arbitrary contexts c, (1) is necessarily false and (2) is contingent. That (2) comes out *a priori*, and (1) merely as epistemically possible is something that should be secured by a proper interpretation of Comp and Bel.

---

[15] Maybe it is best to treat Comp_x, Bel_x, and Has_x as predicates of sentences, in consideration of the metalinguistic nature of their intended interpretation. E.g. x may be a competent speaker of S while failing to be a competent speaker of some S' that is synonymous to S; our practice to ascribe belief *via* disquotation is likewise sensitive to linguistc form, see Kripke (1979). I think conceptions, too, are best individuated in terms of linguistic form rather than content alone (although I need not press the point, here). – Of course this means that a formal semantics for this language would be confronted with the problems posed by the liar and related paradoxes.

[16] This is just the minimal postulate we need to derive consequences from conceivability. Logical closure postulates such as this one owe their validity to the fact that we are dealing with ideal rational agents.

[17] See Kaplan (1989). I only define, in a sketchy form, the semantics of the most important items.

[18] The present definition of logical truth is much closer to Kripke's (1972) definition of analyticity than to Kaplan's, where analyticity is identified with apriority, see Kaplan (1989).

References

Brueckner, Anthony (2001), "Chalmers' conceivability argument for dualism." *Analysis* 61,187-93

Chalmers, David (1996), *The Consious Mind*. Oxford: OUP

Chalmers, David (2002), "Does conceivability entail possibility?" In (T. Gendler and J. Hawthorne, eds.), *Imagination, Conceivability, and Possibility*. Oxford: OUP

Kaplan, David (1989), "Demonstratives". In (Almog et.al. eds.), *Themes from Kaplan*. Oxford:OUP, 481-563

Kirk, Robert (1974),"Zombies vs. Materialists". *Aristotelian-Society Proceedings*. Suppl. 48, 135-152

Kripke, Saul (1972), "Naming and Necessity". In (D. Davidson and G. Harman eds.), *Semantics of Natural Language*, Dordrecht: Reidel, 253-355

Kripke, Saul (1979), "A Puzzle about Belief". In: Margalit, A. (ed.), *Meaning and Use*. Dordrecht: Reidel, 239-83

Kupffer, Manfred: "Kaplan's *A Priori*". Forschungsberichte der Forschergruppe "Logik in der Philosophie". Konstanz 2003

Putnam, Hilary (1975), "The Meaning of 'Meaning'". In (K. Gunderson ed.) *Language, Mind and Knowledge*, Minneapolis: University of Minnesota Press, 131-93

Yablo, Stephen (1993), "Is conceivability a guide to possibility?" *Philosophy and Phenomenological Research*, 53, 1-42

Yablo, Stephen (2000), "Textbook Kripkeanism and the Open Texture of Concepts". *Pacific Philosophical Quarterly* 81, 98-122