

Preprint version.

Published by *Philosophical Quarterly*, Oxford University Press.

For the published version see <http://dx.doi.org/10.1093/pq/pqw066>

The Inconceivable Popularity of Conceivability Arguments

Douglas Campbell, Jack Copeland and Zhuo-Ran Deng

Abstract

Famous examples of conceivability arguments include: (i) Descartes' argument for mind-body dualism; (ii) Kripke's 'modal argument' against psychophysical identity theory; (iii) Chalmers' 'zombie argument' against materialism; and (iv) modal versions of the ontological argument for theism. In this paper we show that for any such conceivability argument, C , there is a corresponding 'mirror argument', M . M is deductively valid and has a conclusion that contradicts C 's conclusion. Hence a proponent of C —henceforth, a 'conceivabilist'—can be warranted in holding that C 's premises are conjointly true only if she can find fault with one of M 's premises. But M 's premises—of which there are just two—are modeled on a pair of C 's premises. The same reasoning that supports the latter supports the former. For this reason a conceivabilist can repudiate M 's premises only on pain of severely undermining C 's premises. We conclude on this basis that all conceivability arguments, including each of (i)—(iv), are fallacious.

1. Introduction

Consider the following familiar situation. Someone alleges that a certain philosophically significant proposition, ϕ , is true. One would dearly like to refute this claim. Moreover, perhaps by way of rubbing salt into the wound, one would like to show, not just that ϕ isn't *actually* true, but that it *can't possibly be true*. That is, one's aim is to prove $\neg\Diamond\phi$. How to proceed?

The obvious method is as follows:

Reductio method: First, prove ϕ is contradictory (i.e., that for some p , $\phi \rightarrow (p \wedge \neg p)$). Second, apply modal logic's Necessitation Rule (which lets $\neg\Diamond\phi$ be derived from $\phi \rightarrow (p \wedge \neg p)$).

Voila! Out pops $\neg\Diamond\phi$. Mission accomplished.

But the *reductio* method has a hitch. Proving that ϕ is contradictory can be, well ... difficult! Sometimes, rack one's brains though one will, no contradiction springs to mind. Perhaps no contradiction is there to be found in ϕ in the first place. How to proceed in such a case? How to proceed, that is, when

one badly wants to prove $\neg\Diamond\phi$ but ϕ seems to be *conceivable*? (By ‘conceivable’ we simply mean ‘consistent’ or ‘does not entail any contradictions’. This species of conceivability corresponds to what Chalmers (2010, 143-148) calls *ideal negative conceivability*.)

Here is an oft-tried approach:

First, identify two other propositions, ψ and μ .

Second, show that ψ is (probably) conceivable, by *trying but failing* to detect a contradiction in it.¹

Third, use the conceivability of ψ to infer $\Diamond\psi$, by applying the general principle that *conceivability entails possibility* (CEP).

Fourth, show that $\Diamond\psi\rightarrow\mu$, and infer μ by *modus ponens*.

Finally, show that $\Diamond\phi\rightarrow\neg\mu$, and infer $\neg\Diamond\phi$ by *modus tollens*.

The form of argument—henceforth the ‘conceivability argument’ (CA)—is as follows. ($\Diamond_c\psi$ represents the claim that ψ is conceivable.)

- | | | |
|-----|---|-----|
| C1. | $\Diamond_c\psi$ | |
| C2. | $\Diamond_c\psi\rightarrow\Diamond\psi$ | CEP |
| C3. | $\Diamond\psi\rightarrow\mu$ | |
| C4. | $\Diamond\phi\rightarrow\neg\mu$ | |
| | | |
| C5. | $\neg\Diamond\phi$ | |

In what follows we begin by reviewing several important arguments of this form. We then show that there is something profoundly rotten in their logic.

2. Examples of CA

If the values of ϕ , ψ , and μ are set as follows,

- ϕ : Pain=C-fibre stimulation
- ψ : Pain \neq C-fibre stimulation
- μ : \square (Pain \neq C-fibre stimulation),

¹ Since (for Gödelean reasons, among others) a contradiction might still be lurking somewhere among ψ 's implications even if our best efforts to find it have so far been unsuccessful, a demonstration that ψ is conceivable will generally be *defeasible*.

then CA becomes the following version of Kripke's (1980) 'modal argument' against psychophysical identity theory:

- K1. $\diamond_e(\text{Pain} \neq \text{C-fibre stimulation})$
 K2. $\diamond_e(\text{Pain} \neq \text{C-fibre stimulation}) \rightarrow \diamond(\text{Pain} \neq \text{C-fibre stimulation})$
 K3. $\diamond(\text{Pain} \neq \text{C-fibre stimulation}) \rightarrow \square(\text{Pain} \neq \text{C-fibre stimulation})$
 K4. $\diamond(\text{Pain} = \text{C-fibre stimulation}) \rightarrow \neg \square(\text{Pain} \neq \text{C-fibre stimulation})$
-
- K5. $\neg \diamond(\text{Pain} = \text{C-fibre stimulation})$

Here K1 is justified by the apparent absence of contradictions in the idea of pain being non-identical to C-fibre stimulation.² K2 is an instance of CEP. K3 is justified by the fact that both 'pain' and 'C-fibre stimulation' are rigid designators, and by Kripke's (1980) principle that all identities and non-identities between rigid designators are necessary. K4 is trivial. K5 is bad news for anyone wishing to identify mental types with physical or functional types, for the argument readily generalizes.

Kripke presents Argument K as a modern take on Descartes' argument for mind-body dualism (1980, pp. 144–150). It is therefore no surprise that Descartes' argument can itself be shoehorned into CA's form. Setting ϕ , ψ , and μ as follows:

- ϕ : Mind = Body
 ψ : Mind \neq Body
 μ : $\square(\text{Mind} \neq \text{Body})$,

we get:

- D1. $\diamond_e(\text{Mind} \neq \text{Body})$
 D2. $\diamond_e(\text{Mind} \neq \text{Body}) \rightarrow \diamond(\text{Mind} \neq \text{Body})$
 D3. $\diamond(\text{Mind} \neq \text{Body}) \rightarrow \square(\text{Mind} \neq \text{Body})$
 D4. $\diamond(\text{Mind} = \text{Body}) \rightarrow \neg \square(\text{Mind} \neq \text{Body})$
-
- D5. $\neg \diamond(\text{Mind} = \text{Body})$

² Kripke frames his argument in terms of 'Pain=C-fibre stimulation' being *a posteriori*, rather than in terms of its denial, 'Pain \neq C-fibre stimulation', being conceivable. But the former implies the latter, since p can be *a posteriori* only if $\neg p$ is contradiction-free, and thus only if $\neg p$ is conceivable.

Argument D appears a plausible rational reconstruction of Descartes' argument in *Meditation VI*:

I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God... [O]n the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of a body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it. (Cottingham, Stoothoff, & Murdoch, 1629, p. 54)

Here Descartes uses a (theistic) version of CEP to infer that it is *possible* for mind and body to be distinct. From this he infers they are *actually* distinct. Why does he think he can make this move from mere possibility to actuality? Presumably because he is assuming D3, or something like it, as a tacit premise (Robinson, 2012).

Chalmers' (1996; 2010) 'zombie argument' against materialism can also be cast as a version of CA. Following Chalmers, let P represent a conjunction of all the microphysical facts. Let T be a 'totality operator' (or 'that's all' clause), which, when tacked onto the end of P , yields a proposition that says P provides a *complete* description of all the non-supervenient facts that obtain in the world. Let Q represent a conjunction of all the phenomenal facts. Thus ' $\Box(PT \rightarrow Q)$ ' represents the materialist thesis that the phenomenal facts supervene metaphysically on the bare microphysical facts (i.e., that any 'PT-world' must also be a 'Q-world').

Plugging the following values for ϕ , ψ , and μ into CA,

$$\begin{aligned}\phi: & PT \wedge Q \\ \psi: & PT \wedge \neg Q \\ \mu: & \neg \Box(PT \rightarrow Q),\end{aligned}$$

we get this version of the zombie argument:

$$\begin{aligned}\text{Z1.} & \Diamond_c(PT \wedge \neg Q) \\ \text{Z2.} & \Diamond_c(PT \wedge \neg Q) \rightarrow \Diamond(PT \wedge \neg Q) \\ \text{Z3.} & \Diamond(PT \wedge \neg Q) \rightarrow \neg \Box(PT \rightarrow Q) \\ \text{Z4.} & \Diamond(PT \wedge Q) \rightarrow \Box(PT \rightarrow Q)\end{aligned}$$

Z5. $\neg\Diamond(P\!T\wedge Q)$.

Here Z1 claims that a PT-world could conceivably fail to be a Q-world. It is justified by the apparent absence of contradictions in the idea of a PT-world being, say, a *zombie world* (a world wherein some or all human beings lack phenomenal consciousness). Z2 is an instance of CEP. Z3 is trivial.

Z1, Z2 and Z3 suffice by themselves to give Chalmers the result he is after, namely $\neg\Box(P\!T\rightarrow Q)$ (the denial of materialism). Chalmers doesn't need Z4, and so Z4 doesn't feature in the zombie argument as Chalmers himself presents it. Hence Chalmers' own version of the zombie argument doesn't quite fit the form of CA.

However, Z4 is a harmless addition to Chalmers' zombie argument because it is *analytic*. Why so? Well, recall that a *PT*-world is a logically possible world where *P* offers a *complete* description of the supervenience base. This being so, any two *PT*-worlds must be alike in all respects. No fact obtaining in one could fail to obtain in the other, every such fact being entailed by *PT*. So if *some* *PT*-world is a *Q*-world (i.e., a world that is a phenomenological duplicate of the actual world), then (by the meaning of the *T* operator) *every* *PT*-world must be a *Q*-world. In short, if $P\!T\wedge Q$ is true at some possible world, then $P\!T\rightarrow Q$ will be true at all possible worlds. This is what Z4 says.

Since Z4 is analytic Chalmers can't object if we add Z4 to his zombie argument as an extra premise, to produce Argument Z, which *does* fit CA's form. If Chalmers' version of the zombie argument is sound, then Argument Z is sound too. Contrawise, if Argument Z is problematic, as we will show below, then so too is Chalmers' zombie argument.

The above examples all come from the philosophy of mind, but CA also crops up elsewhere. A case in point is the modal ontological argument (Hartshorne, 1965; Malcolm, 1960; Plantinga, 1974), of which one formulation, obtained by setting the values of ϕ , ψ and μ as follows,

ϕ : God doesn't exist
 ψ : God exists
 μ : $\Box(\text{God exists})$,

is this:

- O1. $\diamond_c(\text{God exists})$
 O2. $\diamond_c(\text{God exists}) \rightarrow \diamond(\text{God exists})$
 O3. $\diamond(\text{God exists}) \rightarrow \square(\text{God exists})$
 O4. $\diamond(\text{God doesn't exist}) \rightarrow \neg \square(\text{God exists})$
-
- O5. $\neg \diamond(\text{God doesn't exist})$.

Here the justification for O3 turns on the idea that the concept of God is (in part) the concept of a necessarily existent being. The rest of the argument is self-explanatory.

3. Why CA is problematic

Let a ‘conceivabilist’ be a proponent of some version of CA. That is, she is someone who, for certain values of ϕ , ψ , and μ , defends the claim that $\neg \diamond \phi$ is true by arguing that C1, C2, C3 and C4 are true. For example, proponents of Argument K, D, Z and O are conceivabilists.

We now show that the conceivabilist’s position is untenable, and that by relying on CA to argue for $\neg \diamond \phi$ she reasons fallaciously. To see the problem, consider the following ‘mirror argument’:

- M1. $\diamond_c \phi$
 M2. $\diamond_c \phi \rightarrow \diamond \phi$ CEP
-
- $\diamond \phi$

For reasons to be explained in a moment, the arguments the conceivabilist uses to justify two of her own premises, C1 and C2, also justify M1 and M2. But M1 and M2 jointly entail $\diamond \phi$, a conclusion that flatly contradicts $\neg \diamond \phi$, the conclusion she is herself arguing for. And so she is caught in the jaws of an inconsistency.

Why must the conceivabilist accept M1? M1 is modeled on C1: where C1 says ψ is *conceivable*, M1 says ϕ is *conceivable*. The conceivabilist’s reasons for accepting C1, will, if they are any good, consist of the fact that ψ *appears* to be conceivable, in the sense that ψ ’s logical implications appear to be contradiction-free. Now, suppose ϕ *also appeared to be conceivable*—i.e., that its implications also appeared to be contradiction-free. In this case the conceivabilist’s reasons for accepting C1 would be matched by equally good reasons for accepting M1,

and so C1 and M1 would stand or fall together: the conceivabilist could reject M1 only on pain of admitting that C1 might just as easily have been rejected, instead.

This being so, a conceivabilist who opts to reject M1 must first of all break the symmetry between M1 and C1 by showing that whereas ψ appears to be conceivable, ϕ does not. In order to do this she needs to have some argument—call it T —at her disposal, which justifies her in doubting that ϕ is conceivable. T might consist of an outright demonstration of a contradiction in ϕ 's implications, in which case it will provide *absolute certainty* that ϕ is inconceivable. But T needn't be quite so conclusive as this. For instance, it might merely consist of various forceful intuitions to the effect that ϕ is contradictory. If these intuitions have not yet been borne out by the actual detection of the putative contradiction in question, then they won't justify the conceivabilist in being 100% certain that ϕ is contradictory. But they might still justify her in being relatively confident that ϕ is contradictory, and thus relatively confident that ϕ is inconceivable.

Now if the conceivabilist possesses some such T then she can, perhaps, reject M1 without undermining C1 in the process. But here's the rub. In using T to reject M1 the conceivabilist will be protecting CA's premises from being refuted by M1 and M2, but only at the expense of exposing these selfsame premises as being *logically redundant*. To see this, notice that if ϕ entails a contradiction (i.e., if ϕ is inconceivable), then $\neg\Diamond\phi$ can be proved by the *reductio* method (described in §1), instead of by using CA. Hence, just to the degree that T provides the conceivabilist with reason to think that ϕ is contradictory, it also provides her with a direct, *reductio*-method-based proof of $\neg\Diamond\phi$, a proof that is logically independent of CA itself. In other words, just in so far as T justifies the conceivabilist in rejecting M1, it also justifies her in thinking that CA is otiose and dispensable. What T gives the conceivabilist with one hand, by enabling her to defend CA from M1 and M2, it takes back with the other, by rendering CA superfluous.

Compare. A burglar attempts to enter a house through a small window. Finding the window securely latched, he breaks down a door, unlatches the window from the inside, exits through the door, then successfully squeezes back in through the window. Triumph! Of course having broken down the door he no longer needed to bother with entering through the window. Similarly, a conceivabilist who can defend CA's premises from the mirror argument by using T to argue against M1 doesn't need to bother anymore with using CA to prove $\neg\Diamond\phi$. T does the job by itself. CA is surplus to her requirements.

From what has just been said it follows that if CA *is not* logically redundant—i.e., if the conceivabilist is genuinely reliant on CA to prove $\neg\Diamond\phi$ —then she can't have any such argument as *T* at her disposal. She therefore won't be in a position to defend CA's premises from the mirror argument by rejecting M1. Needless to say, if Descartes, Kripke, Chalmers or proponents of the modal ontological argument were able to prove their respective conclusions using the *reductio* method, without relying on CA, then they would be first to recognize and loudly trumpet this fact. They resort to using CA only because they have no such argument as *T* up their sleeves. Hence they are not in a position to repudiate the mirror argument by rejecting M1.

Before we move on, three brief clarifications are in order. The first concerns logical redundancy. Redundancy in one's arguments can, of course, be a useful and desirable thing. If one has two arguments, *G* and *H*, that share the same conclusion, *p*, then *G* can be used as backup in cases where *H* fails to convince, and *vice versa*. Notice, however, that such redundancy is useful only if *G* and *H* are *logically independent* of each other, in the sense that there are good reasons for accepting *G*'s premises that don't presuppose the truth of *H*'s premises, and *vice versa*. If *G*'s premises were vulnerable to some counterargument, and if it were necessary to rely on *H*'s premises in order to defend *G*'s premises from this counterargument, then *G* wouldn't provide any genuine support for *p* over and above the support already provided by *H*. *H* would be doing all the real logical work, and *G* would be otiose.

Unfortunately for the conceivabilist she would, in using *T* to defend CA's premises from the mirror argument, be making CA logically dependent on *T*. The resulting logical redundancy is therefore of the useless variety, not the useful variety. By using *T* to defend CA's premises and then using CA's premises to argue for $\neg\Diamond\phi$ she would be relying on a complete set of premises comprised of all of *T*'s premises and all of CA's premises. But since *T* shows ϕ is contradictory, *T*'s premises suffice by themselves to provide a *reductio*-method-based proof of $\neg\Diamond\phi$. This makes the other premises she is invoking—namely, those that belong to CA but not to *T*—extraneous where the goal of proving $\neg\Diamond\phi$ is concerned. They add nothing but pointless complexity to the overall case for thinking $\neg\Diamond\phi$ is true. They are like cogs in a clockwork that can be removed without disturbing the clockwork's function.

The second clarification concerns the conceivabilist's goal. It is of course *part* of the conceivabilist's goal to prove $\neg\Diamond\phi$. She would succeed in attaining at least this part of her goal were she to use *T* and the *reductio* method to prove $\neg\Diamond\phi$. This is not in dispute. We have no objection to the idea that $\neg\Diamond\phi$ might be

proved using the *reductio* method. But it is the conceivabilist's goal, not to prove $\neg\Diamond\phi$ using *some method or other*, but to prove it *using CA*. It is distinctive of the conceivabilist that she thinks CA is capable of providing us with good reason to accept $\neg\Diamond\phi$. If a conceivabilist were to construct an argument, *T*, with which to attack M1, and furthermore acknowledge that *T* renders CA pointless and superfluous, then we would no longer be in any disagreement with her: but she would have renounced her position and be a 'conceivabilist' no more.

The third clarification concerns *degrees of confidence*. Clearly if *T* provides *absolute certainty* that ϕ is contradictory, then it also enables $\neg\Diamond\phi$ to be proved outright by the *reductio* method, so rendering CA *entirely* redundant. But what if *T* only justifies the conceivabilist in being $x\%$ confident that ϕ is contradictory, where $0 < x < 100$? Might CA have a useful logical role to play in this case? No. To see why not let's distinguish two epistemic possibilities. Possibility 1 is that ϕ is contradictory and M1 is false. Possibility 2 is that ϕ is non-contradictory and M1 is true. *T* enables the conceivabilist to assign a credence of $x\%$ to Possibility 1. This leaves a credence of $(100-x)\%$ to be assigned to Possibility 2 (Possibility 2 being simply the logical complement of Possibility 1). In being $x\%$ confident that Possibility 1 obtains, the conceivabilist can also be $x\%$ confident that CA is logically redundant (because if Possibility 1 obtains then the *reductio* method proves $\neg\Diamond\phi$). In being $(100-x)\%$ confident that Possibility 2 obtains, she can also be $(100-x)\%$ confident that CA's premises cannot be successfully defended from the mirror argument by attacking M1 (because if Possibility 2 obtains, then M1 is true). Putting these two results together, we obtain the conclusion that she can be $x\% + (100-x)\% = 100\%$ confident *that either CA is logically redundant or CA's premises cannot be successfully defended from the mirror argument by attacking M1*. In other words, if there is epistemic uncertainty as to which of these two possibilities obtains then there will be corresponding uncertainty as to which of the two horns of a dilemma the conceivabilist will be impaled by, but this should be of cold comfort to the conceivabilist *because it is still 100% certain that she will be impaled by one horn or the other*.

So much for M1. Next, why can't the conceivabilist reject M2? M2 is modeled on C2, for where C2 says *that if ψ is conceivable then ψ is possible*, M1 says instead *that if ϕ is conceivable then ϕ is possible*. The conceivabilist justifies C2 by invoking CEP, the general principle that if a proposition is conceivable then it is possible. But unfortunately for the conceivabilist this principle justifies M2 every bit as much as it justifies C2. Were the conceivabilist to reject M2 then she would, in effect, be holding that ϕ is

conceivable but impossible. If this were right, then ϕ would be a counterexample to CEP, which would raise the possibility of ψ being another such counterexample. The conceivabilist's own CEP-based argument for C2 would thereby be severely undermined: for if the conceivabilist herself grants that CEP fails where ϕ and M2 are concerned, then why should CEP be trusted where ψ and C2 are concerned?

Summary. On the one hand the conceivabilist uses C1, C2, C3 and C4 to argue for $\neg\Diamond\phi$. On the other hand, her reasons for accepting C1 and C2 also support M1 and M2, which together entail $\Diamond\phi$. Specifically, just as CEP supports C2, so too it supports M2. And just as an apparent absence of contradictions in ψ suggests that C1 is true, so too an apparent absence of contradictions in ϕ would suggest that M1 is true. The conceivabilist is in no position to deny that ϕ appears to be contradiction-free, since in denying this she would be setting up a *reductio*-method-based proof of $\neg\Diamond\phi$, which would render CA logically redundant. The conceivabilist's premises and the principles she uses to justify these premises therefore 'prove too much'. They generate a contradiction, in the form of $\neg\Diamond\phi\wedge\Diamond\phi$.

Other critics of the various different versions of CA have noticed that they are vulnerable to being 'mirrored' along the above lines (although the term 'mirroring' is ours). For example, Bayne (1988) argues that Kripke's argument can be mirrored to yield a conceivability argument *for* (rather than *against*) psychophysical identity theory. Frankish (2007) constructs an 'anti-zombie argument', which amounts to a mirrored version of Chalmers' zombie argument. Marton (1998), Yablo (1999), Sturgeon (2000), and Brown (2010) present similar criticisms of the zombie argument. Where the modal ontological argument is concerned, various authors (see, e.g., McGarth, 1990) have observed that it seems possible to run it backwards, starting from the conceivability of a (necessarily-existent) God *not* existing and then inferring that it is *not possible* for there to be such a God.

However connections are seldom drawn between these disparate literatures. It appears to have gone unrecognized that the logical issues being encountered in each of the cases are, at root, the same, and that the problem is a general one that afflicts all versions of CA identically. Moreover, at least to our minds, none of these authors have exposed the true depth of CA's logical bankruptcy. Their critiques of the various versions of CA suggest the presence of loopholes through which a proponent of CA might escape. (See our discussion of (Zemach 1994) and (Chalmers 2010), below.) We don't think these loopholes are real. By way of showing this we now return to Arguments K, Z,

D and O, and examine how the general mirroring objection we have just outlined plays out in each case.

4. Against Argument K

As explained in §2, Argument K is a version of CA wherein ϕ 's value is 'Pain=C-fibre stimulation'. Plugging this value for ϕ into the mirror argument produces the following mirrored version of Argument K:

$$\begin{array}{l}
 \text{K'1. } \diamond_e(\text{Pain=C-fibre stimulation}) \\
 \text{K'2. } \diamond_e(\text{Pain=C-fibre stimulation}) \rightarrow \diamond(\text{Pain=C-fibre stimulation}) \\
 \hline
 \diamond(\text{Pain=C-fibre stimulation})
 \end{array}$$

K'1 and K'2 conjointly entail $\diamond(\text{Pain=C-fibre stimulation})$, thereby flatly contradicting K1–K4, which conjointly entail $\neg\diamond(\text{Pain=C-fibre stimulation})$. Hence to save his premises Kripke must reject K'1 or K'2.

Can Kripke reject K'1? Suppose he knew of good reasons for thinking that the *Pain=C-fibre stimulation* hypothesis is contradictory. Then he could immediately refute psychophysical identity theory using the *reductio* method, and so Argument K would be redundant. But of course he doesn't use the *reductio* method to prove this result; he relies on Argument K instead. Great philosopher that he is, he wouldn't keep a simple, knockdown, *reductio*-based refutation of psychophysical identity theory secret if he had one up his sleeve. We may therefore conclude that he can't tender good reasons for thinking that the *Pain=C-fibre stimulation* hypothesis is contradictory. This being so, the *Pain=C-fibre stimulation* hypothesis and the *Pain≠C-fibre stimulation* hypothesis are, for Kripke, on a par, in the respect that to the best of his knowledge *neither one* of them entails a contradiction and *both* appear to be conceivable. Kripke uses the apparent conceivability of the *Pain≠C-fibre stimulation* hypothesis to justify K1, which says that *Pain≠C-fibre stimulation* is in fact conceivable.³ By parity of reasoning, the apparent conceivability of the *Pain=C-fibre stimulation* hypothesis likewise justifies K'1, which says that *Pain=C-fibre stimulation* is conceivable. If Kripke is warranted in concluding that the one hypothesis is conceivable based on its appearing to be

³ Kripke (1980) speaks, variously, of *Pain≠C-fibre stimulation* being 'epistemically possible', or of it being '*a posteriori*', or of *Pain=C-fibre stimulation* 'appearing contingent'. He does not, as we do, speak of *Pain≠C-fibre stimulation* being *conceivable*. But this is a mere difference of terminology.

contradiction-free, then he is warranted in concluding that the other hypothesis is conceivable on the same grounds. And so Kripke is obliged to accept K'1.

His only other option is to reject K'2. Kripke (famously) recognizes that there are certain exceptions to CEP, in the form of propositions, like, say, $Water \neq H_2O$, which are conceivable but impossible (or, in his terminology, *a posteriori* but necessarily false). However he notes that such exceptions to CEP involve rigid designators (like the natural-kind term, 'water') that secure reference via accidental properties of their referents. He points out that neither 'Pain' nor 'C-fibre stimulation' is such a rigid designator, since both these terms pick out their referents via *essential* properties—the way pain *feels* in the one case, and the essential scientific nature of C-fibre stimulation in the other. He therefore concludes (albeit somewhat tentatively—see (Kripke 1980, pp. 148 & 150)) that the *Pain ≠ C-fibre stimulation* hypothesis is *not* an exception to CEP. This gives him his premise K2 (which say that if *Pain ≠ C-fibre stimulation* is conceivable then it is possible). But if the *Pain ≠ C-fibre stimulation* hypothesis is free of rigid designators that refer via accidental properties of their referents, then so too is the *Pain = C-fibre stimulation* hypothesis: for the two hypotheses differ only in that one, being the denial of the other, includes an additional negation concept. Hence the restricted version of CEP used by Kripke to justify acceptance of K2 also justifies acceptance of K'2.

In short, although Kripke must reject K'1 or K'2, he can reject them only on pain of admitting that the reasoning he uses to justify K1 and/or K2 cannot be trusted.

This mirroring objection to Kripke's modal argument is partly anticipated by Bayne (1988), who, like us, points out that Kripke's argument is susceptible to being turned on its head.⁴ Bayne's argument has received scant attention, but is critiqued by Zemach (1994), who defends Kripke. Since Zemach's argument against Bayne might be adapted to make trouble for us, it will be instructive to examine it.

Zemach argues, in effect, that there is an asymmetry between Kripke's premise, K2, and the corresponding premise of the mirror argument, K'2.

⁴ There are both terminological and substantive differences between our argument and Bayne's. On the terminological front, Bayne frames his discussion in terms of the 'apparent contingency' of the *Pain ≠ C-fibre stimulation* hypothesis, not, as we do, in terms of the conceivability of the *Pain = C-fibre stimulation* hypothesis. On the substantive front, Bayne doesn't point out, as we do, that Kripke *must* concede that *Pain = C-fibre stimulation* appears conceivable, on pain of rendering his modal argument against psychophysical identity theory redundant. And unlike us Bayne doesn't present his objection to Kripke as being a mere instance of a much more general objection against *all* conceivability arguments.

Specifically, he contends that the latter is vulnerable to a mode of attack against which the former is invulnerable.

Let's start with the vulnerability of K'2. To attack K'2, it would be necessary to show that *pain=C-fibre stimulation* might be conceivable even if it were in fact impossible, which is to say, even if pain and C-fibre stimulation were non-identical. Zemach points out that even if pain and C-fibre stimulation were non-identical, we could still imagine them always co-occurring *as if they were identical*. He holds that we would thereby, in effect, *be imagining them being identical*. This opens the door to rejecting K'2. In arguing along these lines, Zemach is tacitly relying on the following principle:

P: If A and B are two non-identical states, then in order to conceive of A=B being true, it suffices for one to conceive of A occurring whenever B occurs, and *vice versa*.

Next, why according to Zemach is K2 invulnerable to the same style of attack? Well, to attack K2 it would be necessary to show that *pain≠C-fibre stimulation* might be conceivable even if it were in fact impossible, which is to say, even if pain and C-fibre stimulation were in fact identical. But if pain and C-fibre stimulation were identical, then how could one coherently imagine them being non-identical? Zemach notes that this cannot be done by simply imagining that pain and C-fibre stimulation sometimes *fail* to co-occur *as if they were non-identical*, because on the operative assumption that pain and C-fibre stimulation are in fact identical, this is incoherent. (If A and B are identical states, then to imagine A occurring *just is* to imagine B occurring, and *vice versa*.)

That's Zemach's argument. What's wrong with it? Our answer is that P is not remotely credible. Two states can reliably co-occur without being identical, as when they are non-identical states that co-occur by chance alone, or as when they share a common cause. Thus it is simply not the case that conceiving of A and B reliably co-occurring suffices for conceiving of A=B being true.

Zemach might respond by conceding the point, but then switching targets from K'2 to K'1. He might claim that when *we think* we are conceiving of *pain=C-fibre stimulation* we are *really* just conceiving of pain and C-fibre stimulation co-occurring as if they were identical (a kind of mistake that has no analogue where *pain≠C-fibre stimulation* is concerned). He might deny on this basis that *pain=C-fibre stimulation* is genuinely conceivable. In reply we note that: (i) at best Zemach would thereby have explained why K'1 might appear to be true even if it were in fact false. He would not have demonstrated that

pain=C-fibre stimulation entails a contradiction, and so he would not have shown that K'1 is in fact false. And (ii) even if this argument did show that K'1 is false (which it doesn't) then it would thereby save Argument K from being mirrored only at the expense of exposing Argument K as being logically redundant: for, as we have seen, if K'1 is false then Argument K's conclusion can be proved by the *reductio*-method, without using Argument K's premises at all.

5. Against Argument Z

Substituting $PT \wedge Q$ for ϕ within the mirror argument yields the following mirrored version of Argument Z:

$$\begin{array}{l}
 Z'1. \quad \diamond_c(PT \wedge Q) \\
 Z'2. \quad \diamond_c(PT \wedge Q) \rightarrow \diamond(PT \wedge Q) \\
 \hline
 \diamond(PT \wedge Q)
 \end{array}$$

Z'1 and Z'2 together entail $\diamond(PT \wedge Q)$, which contradicts Argument Z's conclusion, $\neg \diamond(PT \wedge Q)$, and therefore also contradicts its premises, Z1—Z4.⁵ To save his premises Chalmers must reject Z'1 or Z'2.

Can Chalmers reject Z'1? Suppose he could show that Z'1 is false by showing $PT \wedge Q$ is contradictory. Having done this he could then use the *reductio* method (of §1) to prove $\neg \diamond(PT \wedge Q)$, from which the denial of materialism can then be derived *via* the following conditional:

$$\text{COND:} \quad \neg \diamond(PT \wedge Q) \rightarrow \neg \Box(PT \rightarrow Q)$$

To see why COND must be accepted, consider a PT-world, w (a possible world that is a minimal physical duplicate of our actual world). (Notice that such a w certainly exists. If materialism is true then w will be identical to the actual world. If materialism is false then w will be a bare physical duplicate of the actual world from which non-physical things have been subtracted.) COND's antecedent says, in effect, that no PT-world is a Q-world. Assume this is true. Then it follows that w is not a Q-world. Thus there is at least one world—namely, w —that is a PT-world but *not* a Q-world. Thus it is not the case that *every* PT-world *is* a Q-world. This is what COND's consequent says.

⁵ $\diamond(PT \wedge Q)$ also entails $\Box(PT \rightarrow Q)$ (i.e., the truth of materialism) via Z4, which was shown to be analytic in §2, above.

And so, assuming COND's antecedent is true, its consequent is true too. Thus, COND itself is true. Q.E.D.⁶

In short, if Chalmers were able to refute Z'1 by showing that $PT \wedge Q$ is contradictory, he could then go on to provide an immediate, slam-dunk, *reductio*- (and COND-) based refutation of materialism. Since Chalmers relies on the zombie argument to refute materialism instead of providing any such *reductio*-based refutation, we can reasonably assume that he is unable to show that $PT \wedge Q$ is contradictory. But this means that for Chalmers $PT \wedge Q$ and $PT \wedge \neg Q$ are on a par, in the respect that to the best of his knowledge neither entails a contradiction, so that they both appear conceivable. His argument for Z1 rests on the apparent conceivability of $PT \wedge \neg Q$. By parity of reasoning, the apparent conceivability of $PT \wedge Q$ supports an argument for Z'1. Hence Chalmers is in no position to reject Z'1.

This means he must instead reject Z'2. Chalmers shares Kripke's reservations about CEP. Like Kripke, he acknowledges that conceivability is not a reliable guide to possibility when rigid designators that secure reference via accidental properties of their referents (e.g., names, natural kind terms and indexicals) are in play.⁷ However he maintains that $PT \wedge \neg Q$ is free of such rigid designators, and thus that conceivability entails possibility at least where $PT \wedge \neg Q$ is concerned. This gives him his premise, Z2. But if $PT \wedge \neg Q$ is free of such rigid designators, then so too is $PT \wedge Q$, since the latter proposition differs from the former only in respect of containing one less negation concept. Hence by Chalmers' own reasoning we can conclude that if $PT \wedge Q$ is conceivable then it is possible. This gives us Z'2.

And so Chalmers is caught in the same trap as Kripke. He must reject Z'1 or Z'2, but can do this only by admitting that there is something wrong in his own arguments for Z1 and Z2.

Several authors—e.g., Marton (1998), Yablo (1999), Sturgeon (2000), Frankish (2007), and Brown (2010)—have argued, similarly to us, that the zombie argument's logic can be hijacked to produce a conclusion inconsistent with its own conclusion. Chalmers has two main counter-arguments. First he notes that whereas his own Argument Z trades on the conceivability of a non-modal claim—namely, $PT \wedge \neg Q$ —Marton, Yablo and Sturgeon instead rely on the conceivability of a modal claim—namely, $\Box(PT \rightarrow Q)$ (materialism). He then points out that: (i) there is room for him to deny that CEP applies to such modal

⁶ COND and Z4 together entail the bi-conditional, $\Diamond(PT \wedge Q) \leftrightarrow \Box(PT \rightarrow Q)$ (i.e., $\Diamond(PT \wedge Q) \leftrightarrow$ materialism).

⁷ In Chalmers' (1996, 2010) terminology, conceivability is an unreliable guide to possibility when the primary and secondary intensions of a proposition diverge.

claims; and (ii) such modal claims concern the constitution of the entire space of possible worlds, not just the constitution of a single possible world, which makes their conceivability difficult to evaluate (Chalmers 2010, 179-180). This counter-argument is of no help to Chalmers where Argument *Z'* is concerned, since Argument *Z'* is just like Argument *Z* in that it trades on the conceivability of a non-modal claim—namely, $PT \wedge Q$.

Chalmers' second counter-argument is as follows:

It may be prima facie negatively conceivable that materialism is true about consciousness, but it is not obviously conceivable in any stronger sense. Many people have noted that it is very hard to imagine that consciousness is a physical process. I do not think that this unimaginability is so obvious that it should be used as a *premise* in an argument against materialism, but likewise the imaginability claim cannot be used as a premise either. (2010, p. 180)

Chalmers is here attempting to negotiate a safe path between two horns of a dilemma. On the one hand he doesn't want to grant that materialism is manifestly *inconceivable* (i.e., that it clearly entails a contradiction) because then materialism could be refuted by a simple *reductio* argument and so his zombie argument against materialism would be superfluous. Call this *the threat of redundancy*. On the other hand, he mustn't grant that materialism is *conceivable* either, because then CEP could be used to infer that materialism is *possible*, and from this it would follow (via the fact that materialism is itself a modal claim) that materialism is *true*. Call this *the threat of absurdity*.

Chalmers' way out of the dilemma, as intimated in the above passage, is to maintain that it is *epistemically uncertain* whether materialism is conceivable or inconceivable. If Chalmers is right then: (i) uncertainty about whether materialism is inconceivable prevents materialism being proved false with a *reductio* argument and so saves the zombie argument from the threat of redundancy; and (ii) uncertainty about whether materialism is conceivable prevents the zombie argument's logic from being hijacked to prove materialism true, thereby saving it from the threat of absurdity.

That's Chalmers' general strategy. He could bring it to bear against our mirror argument by maintaining that it is epistemically uncertain whether $PT \wedge Q$ is conceivable or inconceivable, and thus epistemically uncertain whether *Z'1* is true or false. But this approach fails for reasons previewed in §3, above. Viz., uncertainty about whether *Z'1* is true or false merely translates into uncertainty about *which* horn of the dilemma Chalmers will be skewered by, not into uncertainty as to *whether* he will be skewered by one horn or the other. To see this, suppose Chalmers has reasons for being $x\%$ confident that $PT \wedge Q$ is

inconceivable, and thus for being $(100-x)\%$ confident that $PT\wedge Q$ is conceivable. Suppose, furthermore, that these reasons leave him epistemically uncertain whether $PT\wedge Q$ is conceivable or inconceivable. Thus $0\ll x\ll 1$. Given Chalmers is $x\%$ confident that $PT\wedge Q$ is inconceivable, he should also be $x\%$ confident that $\neg\Diamond(PT\wedge Q)$ can be proved by the *reductio* method, instead of by using CA, and so he should be $x\%$ confident that Argument Z succumbs to the threat of redundancy. Furthermore, given he is $(100-x)\%$ confident that $PT\wedge Q$ is conceivable, he should be $(100-x)\%$ confident that Argument Z succumbs to the threat of absurdity; for if $PT\wedge Q$ is conceivable then Z'1 is true, and if Z'1 is true then Z'1 and Z'2 together entail $\Diamond(PT\wedge Q)$, which contradicts Argument's Z's conclusion. Putting these two results together, the credence he assigns to the proposition that Argument Z succumbs *either* to the threat of absurdity *or* to the threat of redundancy should be $x\%+(100-x)\%=100\%$.

6. Against Arguments D and O

We will keep our comments on Arguments D and O brief.

Substituting *Mind=Body* for ϕ in the mirror argument yields this mirrored version of Argument D:

$$\begin{array}{l} D'1. \quad \Diamond_e(\text{Mind}=\text{Body}) \\ D'2. \quad \Diamond_e(\text{Mind}=\text{Body}) \rightarrow \Diamond(\text{Mind}=\text{Body}) \\ \hline \Diamond(\text{Mind}=\text{Body}) \end{array}$$

Similarly, substituting *God doesn't exist* for ϕ in the mirror argument yields the following mirrored form of Argument O:

$$\begin{array}{l} O'1. \quad \Diamond_e(\text{God doesn't exist}) \\ O'2. \quad \Diamond_e(\text{God doesn't exist}) \rightarrow \Diamond(\text{God doesn't exist}) \\ \hline \Diamond(\text{God doesn't exist}) \end{array}$$

D'1 and D'2 conjointly entail $\Diamond(\text{Mind}=\text{Body})$, thereby contradicting Argument D's conclusion. Hence a Cartesian proponent of Argument D must repudiate D'1 or D'2. By the same token a theist proponent of Argument O must repudiate O'1 or O'2.

Might the Cartesian reject D'1, or might the theist reject O'1? If the Cartesian could refute D'1 by demonstrating the existence of a contradiction in

the *Mind=Body* hypothesis then she could go on to offer a simple, knockdown proof of $\neg\Diamond(Mind=Body)$ using the *reductio* method, and so Argument D would be logically redundant. Since she relies on Argument D to prove $\neg\Diamond(Mind=Body)$ instead of using a *reductio* argument, she presumably knows of no such contradiction. This being so, the *Mind \neq Body* and *Mind=Body* hypotheses will be on a par for the Cartesian, in the respect that both appear to be contradiction-free. Her grounds for accepting the conceivability of the latter hypothesis are therefore just as good as her grounds for accepting the conceivability of the former. This means she is obliged to accept D'1 for the same reasons she accepts D1. Were she to fail to accept D'1 she would be tacitly acknowledging a weakness in her argument for D1.

What has just been said about the Cartesian holds equally for the theist. If the theist could show that *God doesn't exist* entails a contradiction, then she could prove God's existence by *reductio*, and so she wouldn't need Argument O in the first place. Given that she instead relies on Argument O, we can infer that she can't show that *God doesn't exist* entails a contradiction. But this being so, *God exists* and *God doesn't exist* are for the conceivabilist on a par: they both appear conceivable. The same considerations that drive her to accept O1 should also drive her to accept O'1. She could balk at accepting O'1 only at the cost of acknowledging a weakness in her argument for O1.

If the Cartesian doesn't reject D'1, she must instead reject D'2. Likewise, if the theist does reject O'1, she must instead reject O'2. But both D'2 and O'2 are instances of CEP. If D'2 is false, it is a counterexample to CEP; and likewise for O'2. Neither the conceivabilist nor the theist can afford to concede that CEP has such counterexamples, because the conceivabilist relies on CEP to justify D2, and the theist relies on it to justify O2.

Thus the Cartesian and the theist are caught in the same trap as Kripke and Chalmers: they must reject one of the mirror argument's premises in order to avoid an outright contradiction, but in rejecting either one of the mirror argument's premises they would be conceding that their own premises are inadequately supported.

7. Conclusion

We began with the question as to how one might prove $\neg\Diamond\phi$ if, due to the fact that ϕ appears to be contradiction-free, one is unable to prove it by the *reductio* method. The conceivabilist thinks CA provides an answer. We hope to have persuaded the reader that it does not. The problem is really perfectly simple. If ϕ appears to be contradiction-free, then ϕ appears to be conceivable,

and if ϕ appears to be conceivable then the conceivabilist's own logic can be easily adapted to prove $\diamond\phi$, a conclusion that exactly contradicts the conclusion the conceivabilist is aiming for! This adapted version of CA, which proves $\diamond\phi$, is our 'mirror argument'.

We have shown that four famous arguments—namely, Kripke's modal argument, the Cartesian argument for dualism, Chalmers' zombie argument and the modal ontological argument—can each be cast as versions of CA, and that each are susceptible to being mirrored. We conclude that all four of these arguments are logically bankrupt.

Acknowledgements

Thanks to Michael-John Turp and Diane Proudfoot for helpful comments.

References

Bayne, S. R. (1988). Kripke's Cartesian Argument. *Philosophia*, 18(July),

265–270.

Brown, R. (2010). Deprioritizing the A Priori Arguments Against Physicalism.

Journal of Consciousness Studies, 17(3-4), 47–69.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental*

Theory (Vol. 4). Oxford University Press.

Chalmers, D. J. (2010). *The Character of Consciousness*. Oxford University

Press.

Cottingham, J., Stoothoff, R., & Murdoch, D. (1629). *The Philosophical*

Writings of Descartes. Cambridge University Press.

Frankish, K. (2007). The Anti-Zombie Argument. *Philosophical Quarterly*,

57(229), 650–666.

- Hartshorne, C. (1965). *Anselm's Discovery: A Re-Examination of the Ontological Proof for God's Existence*. La Salle, IL: Open Court.
- Kripke, S. A. (1980). *Naming and Necessity* (Vol. 217). Harvard University Press.
- Malcolm, N. (1960). Anselm's Ontological Arguments. *Philosophical Review*, 69, 41–62.
- Marton, P. (1998). Zombies Versus Materialists: The Battle for Conceivability. *Southwest Philosophy Review*, 14(1), 131–138.
- McGarth, P. J. (1990). The Refutation of the Ontological Argument. *Philosophical Quarterly*, 40(159), 195–212.
- Plantinga, A. (1974). *The Nature of Necessity*. Oxford: Oxford University Press.
- Robinson, H. (2012). Dualism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012). Retrieved from <http://plato.stanford.edu/archives/win2012/entries/dualism/>
- Sturgeon, S. (2000). *Matters of Mind: Consciousness, Reason and Nature* (Vol. 111). Routledge.
- Yablo, S. (1999). Concepts and Consciousness. *Philosophy and Phenomenological Research*, 59(2), 455–463.