Realism and the logic of conceivability

Dominik Kauss

Published online: 30 January 2020
© Springer Nature B.V. 2020

Abstract On their alethic reading, formulas (T), (D), and (K) codify three of the most basic principles of possibility and its dual (necessity). This paper discusses these formulas on a broadly epistemic reading, and in particular as candidate principles about conceivability and its dual (inconceivability of the opposite). As will be shown, the question whether (T) and its classical dual equivalent, as well as (D) and (K) hold on this reading is not only a logical one but involves a distinctively metaphysical controversy between realist and antirealist views on the relation between truth on the one hand and various cognitive conditions such as knowability, conceivability, and thinkability on the other. It will be argued that the stance we take with regard to the metaphysical dispute has consequences for our assessment of the plausibility not only of (T) and its classical equivalent, but also of (D) and—when that stance is combined with a structural account of propositions—potentially of (K) as well; with all four taken in the above epistemic sense. A second upshot will be that the same sensitivity to metaphysical background commitment also applies to our view as to whether or not inconceivability of the opposite coincides with, or even entails, apriority.

Keywords Conceivability · Thinkability · Apriority · Epistemic modality · Realism · Non-normal logics

Dominik Kauss
kauss@em.uni-frankfurt.de

1 Goethe University Frankfurt, Norbert Wollheim Platz 1, 60629 Frankfurt am Main, Germany
2 Massachusetts Institute of Technology, 77 Massachusetts Avenue, 32-d808, Cambridge, MA 02139-4307, USA
1 Introduction

If you’re interested in the logic of conceivability, chances are that at some point you’ll be wondering which of the principles that we know to hold for alethic possibility also hold for conceivability, and what kind of philosophical commitment one incurs by accepting any of those principles for conceivability. Among the most basic platitudes shaping our notion of alethic modality are the duality principle, stating that necessity equals impossibility of the opposite: $\Box P \iff \neg \Diamond \neg P$; the (T) principle, stating that necessity entails truth: $\Box P \rightarrow P$; the (D) principle, stating that necessity entails possibility: $\Box P \rightarrow \Diamond P$; and the (K) principle, stating that necessity distributes over the conditional: $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$. Here, $\Box$ and $\Diamond$ abbreviate ‘it’s necessary that’ and ‘it’s possible that’, respectively, pertaining not to what is necessary or possible relative to our knowledge or imaginative capacities, say, but to what is necessary or possible tout court. Deviating from this alethic interpretation, we will discuss (T), (D), and (K) on a broadly epistemic reading—call it the conceivability interpretation—which I will flag by painting box and diamond black. In a nutshell, we will use $\Diamond$ to abbreviate ‘it’s conceivable that’, defining $\Box$ as short for $\neg \Diamond \neg$, thus validating the corresponding duality principle $\Box P \iff \neg \Diamond \neg P$ by fiat. As far as (T), (D), and (K) are concerned, we shall see that on the conceivability interpretation they are not quite as immune to objection as they are on the alethic interpretation.1

Ever since Kripke (1980), philosophers have become accustomed to distinguishing the alethic reading of ‘possible’ and ‘necessary’ from a so-called epistemic reading, or a family of broadly epistemic readings. And of course the notion of conceivability has figured prominently in various attempts to explicate the epistemic reading. (See Gendler and Hawthorne 2002.) With a view to the circumstance that the dual of possibility in the alethic sense is necessity in a corresponding sense, it’s natural to expect that the dual of conceivability will likewise be tantamount to some brand of necessity, presumably in a broadly epistemic sense. For instance, you might expect that dual to be equivalent to apriority, or to some notion entailing apriority, say Kripkean analyticity. (See Kripke 1980: 39.) However, whether or not conceivability’s dual indeed is or even entails apriority will turn out to be a substantial question further down the road. In order not to prejudice the issue from the start, we will refer to that dual more neutrally as inconceivability of the opposite.2

Syntactically, the formulae displayed in the first paragraph admit of at least two readings: the schematic reading, on which the letters ‘P’ and ‘Q’ stand for arbitrary unspecified sentences of some object-language; and the quantified reading, on which those letters are understood as variables bound by implicit second-order

---

1 Here and in the following, by ‘interpretation’ I don’t mean anything more technical than simply the vernacular counterparts which box and diamond are used to abbreviate in a given application of modal logic.

2 I will use ‘it’s inconceivable that $P$’ as synonymous with ‘it’s not conceivable that $P$’. This use differs from other legitimate uses, such as Yablo’s (1993: 11f, 29ff.), on which the former phrase is strictly stronger by way of entailing without being entailed by the latter.
universal quantifiers. In Sects. 2–4 we will be interested in the quantified reading of
the conceivability-versions of (T), (D), (K) and similarly ambiguous formulae (at
least outside of subproofs in which quantifiers are temporarily eliminated and ‘\(P\)
and ‘\(Q\)’ occur as free variables). In Sect. 5, I will briefly comment on the extent to
which the observations made in those sections have consequences for the
conceivability versions of (T), (D), and (K) on the schematic reading.

While there are several ways to understand conceivability talk, in the following
only two notions of conceivability will be of relevance. In the first sense, it’s
conceivable that \(P\) iff it’s rationally believable that \(P\); or, more generally, iff it’s
rationally feasible to adopt some non-zero credence (“degree of belief”) in \(P\);
rationally feasible, that is, given suitable evidence, and in the light of ideal rational
reflection. For definiteness, call this doxastic conceivability.

In the second relevant sense, it’s conceivable that \(P\) iff it’s believable (or, more
generally, iff it’s rationally feasible to adopt a non-zero credence) that it’s possible
that \(P\); again, given suitable evidence, and in the light of ideal rational reflection. At
least paradigmatically, cases where it’s conceivable that \(P\) in this second sense are
cases where it’s rationally feasible to imagine a scenario in which \(P\). For this reason,
as well as for the sake of definiteness, let’s call this second concept imaginative
conceivability.3

Obviously enough, doxastic conceivability entails imaginative conceivability: If I
even have a positive credence that \(P\), then I can certainly have a positive
credence that possibly \(P\). The two notions are nevertheless distinct, as the converse
entailment fails: It’s imaginatively conceivable that someone other than the actual
inventor of the zip invented the zip, although this is not doxastically conceivable.4
When in the following I speak of conceivability/\(\) without further qualification, I
will be speaking generally so as to cover both doxastic and imaginative
conceivability.

Before entering the discussion, let’s compare our two conceivability concepts to
more or less similar concepts from the literature. Chalmers distinguishes between
prima facie vs. ideal conceivability, primary vs. secondary conceivability, and
negative vs. positive conceivability. As defined, both doxastic and imaginative
conceivability are ideal, rather than prima facie. Moreover, at least extensionally,
the two seem roughly analogous to Chalmers’ distinction between primary and
secondary conceivability (in this order), as suggested by the fact that it’s secondarily
but not primarly conceivable that someone other than the actual inventor of the zip
invented the zip. It’s more difficult to say at this point whether doxastic and
imaginative conceivability would have to be classified as negative or as positive
notions in Chalmers’ sense. According to Chalmers (2002: 149f.), it’s negatively
(ideally, primarily) conceivable that \(P\) when it’s not ruled out a priori that \(P\), or
when there is no (apparent) contradiction in \(P\); whereas positive notions of

---

3 As words are being used here, in order for it to be rationally feasible to imagine a scenario in which \(P\),
there is no requirement that it be feasible to visually imagine such a scenario. For instance, in the intended
general sense, we can imagine a scenario featuring an invisible man performing such and such deeds,
even though it may not be feasible to visually imagine such a scenario.

4 The example is of course Evans’ (1979), although he uses it to illustrate a slightly different claim.
conceivability require that one can form some sort of positive conception of a situation in which it’s the case that \( P \). Chalmers (2002: 156) concedes that negative conceivability is at least better defined than positive conceivability. And indeed, it’s far from obvious what it means to have a positive conception, especially if you wish to keep the notion of positive conceivability more general than that of visual imaginability. On the other hand, saying that doxastic/imaginative conceivability is negative in Chalmers’ sense would be to prejudge precisely the question we are trying to remain neutral about, viz. whether or not the dual of doxastic conceivability is apriority, and whether or not the dual of imaginative conceivability entails apriority. I’ll return to this point in Sects. 2 and 6.

Among the six notions of conceivability that Yablo (1993) discerns, those that come closest to doxastic and imaginative conceivability, at least superficially, are what he calls conceivability \( \text{qua} \) believability and conceivability \( \text{qua} \) believability of possibility, respectively. However, it must be stressed that the notion of believability he’s working with (which he borrows from the tradition via Reid, Kneale, and others) is rather negative, requiring no more than the inability to rule out the proposition in question. (See Yablo 1993: 7.) One of the conclusions that one might draw from the pending considerations is that such inability may not be sufficient for believability on any ultimately satisfactory analysis. For this reason, I will not assume that our distinction is equivalent to Yablo’s.

With these preliminaries out of the way, let’s now go through (T), (D), and (K) on a conceivability interpretation and examine their respective plausibility.

2 The (T) formula

On its conceivability interpretation,

\[(T) \quad \Box P \rightarrow P\]

says that if it’s inconceivable that not-\( P \), then \( P \); in effect claiming the factivity of the notion of inconceivability of the opposite. While (T) is not obviously false on this interpretation, it’s not quite as trivial, to say the least, as it is on its alethic interpretation. So one might legitimately ask what reason one could have to accept (T) for conceivability.

One such reason, at least \textit{prima facie}, would be the idea that it’s (doxastically) conceivable that \( P \) iff it’s not a priori that not-\( P \), paired with a factive reading of ‘it’s a priori that \( P \). If you think that (doxastic) conceivability and apriority relate to one another as duals in this way, you should also think that

5 The claim that ‘it’s a priori that \( P \) is factive may not be obvious, especially if that locution is read as short for ‘it’s a priori justifiable to believe that \( P \). For a discussion of fallibilist accounts of a priori justification, see Casullo (2003: 56ff.) However, here ‘it’s a priori that \( P \) is understood as short for ‘it’s a priori knowable that \( P \). While many philosophers have assumed knowability to be non-factive—e.g. Williamson (2000: 292)—Brogaard and Salerno (2008) and Fuhrmann (2014) argue quite plausibly that our ordinary notion of knowability is factive. If so, apriority in the intended sense is factive due to the factivity of knowability.
it’s (doxastically) inconceivable that not-$P$ iff it’s a priori that $P$;

and given that apriority is factive, by transitivity you should think that if it’s unconceivable that not-$P$ then $P$, which is just what (T) claims. However, later in this section we will review an argument that can be used against the idea that the dual of conceivability is, or even entails, apriority.

Note that, given classical propositional logic and the definitional duality of $\Box$ and $\Diamond$, (T) is equivalent to the claim that truth entails conceivability:

$$(T') \quad P \rightarrow \Diamond P$$

One reason for accepting $(T')$ would be one’s prior commitment to an antirealist view of truth. Antirealists maintain that truth entails knowability, or rational believability; be this because they directly analyze or explicate the notion of truth in terms of some cognitive condition (such as verifiability, warranted assertibility, scientific agreement at the end of enquiry, or what have you) or because they have some less direct argument to this conclusion. (See Peirce 1878; James 1909; Wittgenstein 1975: 148; Strawson 1966; Dummett 1976: 100; Putnam 1981; Hofweber 2019.) Of course, knowability/rational believability entails doxastic conceivability: If one can even know/rationally believe that $P$, then one can also assign $P$ a credence above 0. And as already noted, if one can do that much, then a fortiori one can do the same with the proposition that it’s possible that $P$. Hence, if indeed every truth is knowable/rationally believable, as the antirealist has it, then, by transitivity, every truth is conceivable (in either sense), which is just what $(T')$ says.

Meanwhile, this particular reason for accepting $(T')$ does not automatically convert into a reason for accepting (T). While antirealists take truth to be epistemically constrained, many of them do not wish to rule out that there are propositions $P$ such that neither $P$ nor not-$P$ is knowable/rationally believable; propositions, that is, which by antirealist lights are neither true nor false. Combining the former commitment with the latter stance of open-mindedness requires rejecting the law of excluded middle as well as double negation, which antirealists are happy to do. And absent the latter rule, the classical derivation of (T) from $(T')$ is blocked. That said, if you are an antirealist, at least $(T')$ should strike you as highly congenial on the intended interpretation.6

But then of course antirealism itself is highly controversial. Realists deny that there’s an a priori argument to be had that truth entails knowability, rational believability, or indeed any such cognitive -ability condition. (See Nagel 1986: 90ff.; Sorensen 1988: 84, 117ff.; Williamson 2000, 2007.) Accordingly, the most general realist arguments against the antirealist’s knowability/believability claim also tell against $(T')$. Moreover, since realists do not think that truth is epistemically

6 Antirealists are known for their resistance to interdefining box and diamond as duals. Presumably, this resistance partially flows from their having in mind a certain preformal interpretation of box and diamond as necessity and possibility, taken in some epistemic sense. But recall that our abbreviative definition of $\Box$ as inconceivability of the opposite does not presuppose this notion to be equivalent to any sort of necessity, epistemic or alethic, so antirealists should have no reservations about accepting that definition.
constrained, they lack the only relevant reason not to accept double negation, which extends their negative commitment regarding \((T')\) to \((T)\) itself.

Let’s take a look at one such realist argument. In the context of criticizing McDowell’s (1994) claim of the unboundedness of the conceptual, Williamson (2007: 16f.) challenges the underlying assumption that “any object can be thought of,” and, “[i]kewise for the sort of thing that can be the case […], that whenever an object has a property, it can be thought, of the object and the property, that the former has the latter;” in effect: “that everything (object, property, relation, state of affairs, …) is thinkable.” Williamson writes:

What reason have we to assume that reality does not contain elusive objects, incapable in principle of being individually thought of? Although we can think of them collectively—for example, as elusive objects—that is not to single out any one of them in thought. Can we be sure that ordinary material objects do not consist of clouds of elusive sub-sub-atomic particles? We might know them by their collective effects while unable to think of any single one of them. The general question whether there can be elusive objects looks like a good candidate for philosophical consideration. Of course, McDowell does not intend the conceptual to be limited by the merely medical limitations of human beings, but the elusiveness may run deeper than that: the nature of the objects may preclude the kind of separable causal interaction with complex beings that isolating them in thought would require. […] Although elusive objects belong to the same very general ontological category of objects as those we can single out, their possibility still undermines McDowell’s claim that we cannot make “interesting sense” of the idea of something outside the conceptual realm (1994, pp. 105–106). We do not know whether there actually are elusive objects. What would motivate the claim that there are none, if not some form of idealism very far from McDowell’s intentions? We should adopt no […] philosophy that on methodological grounds excludes elusive objects.

How does the notion of thinkability at work here relate to conceivability? Think of thinkability as the cognitive analogue to the linguistic notion of expressibility. For heuristic purposes, it may be best to approach this concept in an agent-relative and circumstantial sense first. We may say that proposition \(P\) is thinkable for agent \(x\) iff \(x\)’s conceptual repertoire and referential abilities allow \(x\) to think \(P\), or as it were to process \(P\) in \(x\)’s “language of thought.” We can then say that in an absolute sense—which is the one we will be concerned with—\(P\) is thinkable iff it’s feasible in principle to acquire conceptual and referential capacities that would enable one to think \(P\). Given this gloss, it should be clear enough that conceivability entails thinkability, as do knowability, believability etc. Since this is the first of a number of background principles that we will avail ourselves of, let’s give it a name and put in on display for further reference:

\[(P1) \quad \Diamond P \rightarrow \Box P\]
Here, ▲ abbreviates ‘it’s thinkable that’. To illustrate the non-equivalence between conceivability and thinkability, note that (P1)’s converse does not hold: For instance, it’s thinkable but not conceivable that one plus one is three.

Now, in what way does Williamson’s Elusive Objects Hypothesis create a problem for antirealism? Well, if there are elusive objects, then for any such object \( x \) there will be at least one property \( F \) such that it’s true that \( x \) has \( F \), even though this very proposition—that \( x \) has \( F \)—is not thinkable; the reason being that thinking it would require the ability to singularly refer to \( x \), which by hypothesis is excluded. In other words, if there are elusive objects, then there are unthinkable truths, and hence truths that fail to be knowable, rationally believable etc., contrary to antirealism.\(^7\)

It might be worthwhile to caution against one way of prematurely dismissing the considered hypothesis out of hand. One may fancy that the idea of there being something that cannot be referred to is a non-starter for the reason that for any term \( t \) and any object \( o \) in a respective domain, there is an “interpretation function” from terms into objects that yields \( o \) as output when given \( t \) as input. Functions of this kind are the bread and butter of standard formal semantics, and one thing they allow us to do is to reconstruct in set-theoretic terms the difference between, on the one hand, sentences whose truth-value depends i.a. on the reference of their embedded singular terms (i.e. they are true with respect to some but not all interpretation function/domain pairs) and, on the other hand, sentences whose truth-value is independent of the reference of their embedded singular terms (i.e. they are true with respect to all such pairs). In reply to this objection let it be conceded that, of course, any object can be the value of a function from terms into objects, and in particular this would hold for elusive objects, if such there be. So, in this superficial sense, any object can be “referred to.” However, this concession does not take the punch out of the Elusive Objects Hypothesis, which addresses a more substantial notion of reference than that. Speaking in terms of the substantial notion, a name refers to whatever object has been given the name; a demonstrative refers to whatever object is demonstrated in the context of utterance; and a definite description refers to whatever object uniquely satisfies it. What the Elusive Objects Hypothesis calls into question is whether everything can be given a name, or be demonstrated, or be uniquely described; not whether for each term and object there is a function that yields the latter as output when taking the former as input.

Above I anticipated that accepting the considered realist argument commits us to rejecting (T). To see this, assume that there are unthinkable truths:

\(^7\) It is not uncommon to see philosophers define or introduce the term ‘proposition’ in terms of something along the lines of ‘potential content of thought’. This definition makes it trivially false that there are or may be unthinkable propositions, and thus is unacceptable to realists who want to claim the latter. Moreover, paired with the idea/heuristic that quantification into sentence position is tantamount to quantification over propositions, that definition is bound to strike the realist as having the effect of restricting the domain of such quantification to the thinkable. Faced with this complication, realists may either reject the definition or rephrase their point, as well as the mentioned idea/heuristic about sentential quantification, in terms of ‘state of affairs’ in place of ‘proposition’. That said, I will continue to use ‘proposition’, without of course assuming the above definition.
If (1) is true, then, since unthinkability entails inconceivability, both doxastic and imaginative, the following is true as well, on either reading:

(2) \( \exists P(P \land \neg \Diamond P) \)

Given classical logic, (2) is equivalent to the negation of (universally quantified) \((T')\). To be sure, deriving that negation from (2) involves relying on rules which many antirealists reject; in particular, the duality of the quantifiers and double negation. As a consequence, pointing out this entailment would be dialectically idle \textit{vis-à-vis} an antirealist who rejects classical logic. However, it’s the realist’s theoretical commitments that we are concerned with now, and the realist accepts classical logic.

Of course, Williamson does not claim (1). Rather, what he claims is that we do not know that (1) is false. And this negative knowledge claim seems to imply, at least pragmatically, that we can and should assign (1) a credence above 0, to the effect that it’s doxastically conceivable that there are unthinkable truths:

(3) \( \Diamond \exists P(P \land \neg \Box P) \)

Since doxastic conceivability entails imaginative conceivability, \( \Diamond \) may be read in either sense in (3). Now, if the logic of conceivability is at least as strong as K, the weakest normal modal logic, then the realist’s implied conceivability claim (3) entails that (2) is conceivably true, i.e. that \textit{it’s conceivable that there are inconceivable truths}, again in either sense:

(4) \( \Diamond \exists P(P \land \neg \Diamond P) \)

One of the nice features of system K is that, when \( Q \) is derivable from \( P \) in some base logic, then diamond-\( Q \) is derivable from diamond-\( P \) in K mounted on top of that base logic. Since (1) entails (2), (2) should be derivable from (1) in any adequate logic of conceivability and thinkability. And since (3) and (4) are simply the \( \Diamond \)-weakenings of (1) and (2), (4) should be derivable from (3) provided that the dual of conceivability satisfies the two characteristic ingredients of K, these being the rule of necessitation and the \((K)\) formula.

In conditional form, necessitation states that if \( P \) is derivable, so is \( \Box P \). \footnote{For simplicity, I’m ignoring the fact that the rule is standardly put in terms of sentences rather than in terms of the propositions expressed by those sentences.} In an alethic setting, this corresponds to the idea that logical truths are necessarily true. Plausible as this may be in a vast number of cases, the idea’s general validity may be called into question with regard to propositions expressible through sentences embedding two-dimensional operators, such as ‘actually’. Thus, one might argue that it’s logically true that if \( x \) invented the zip, \( x \) actually invented the zip; even though it’s possible that someone invented the zip who didn’t actually invent the zip. However, instead of taking this observation to be grounds for simply rejecting necessitation, we may preserve what was initially plausible about the latter by restricting it to derivability in a one-dimensional language.
Transplanted into a conceivability setting, necessitation corresponds to the idea that if \( P \) is derivable, then \( P \)'s negation isn’t even conceivable; at least not *ultima facie*, not in the light of ideal rational reflection. In other words, if \( P \) is derivable, you may as well go on and conclude \( \Box P \). While this is plausible for doxastic conceivability, on an imaginative reading it invites an objection analogous to the one above: It’s logically true that if \( x \) invented the zip, \( x \) actually invented the zip; even though the negation of this truth is imaginatively conceivable. Thus, as in the alethic case, the equivalent of necessitation for imaginative conceivability should be restricted to one-dimensional derivability.

Given necessitation for conceivability, (4) in effect denies that (T)/(T') is a logical truth. Thus, given \( K \) and a classical base logic, the realist’s mere conceivability claim (3) is inconsistent with (T)/(T') for conceivability. The antirealist on the other hand is committed to rejecting (4) as being just as absurd as the corresponding possibility claim, viz. that it’s possible that there are impossible truths. But although (4) *would* be obviously false on its alethic interpretation, it is *not* obviously false on a conceivability interpretation.

As noted, besides \( \Box \) necessitation, our simple argument for the incompatibility between realism and (T)/(T') also relies on (K), to be discussed in Sect. 4. In Sect. 5 the argument will be retroactively qualified by restricting (K) in response to an objection derived from a special version of realism. But this qualification will not hamper the overall objective of this paper, which is to highlight the sensitivity to metaphysical background commitment which (T)/(T'), (D), and (K) bring with them on a conceivability interpretation.

3 The (D) formula

On its conceivability interpretation,

\[
\text{(D) } \Box P \rightarrow \Diamond P
\]

says that if it’s inconceivable that not-\( P \), then it’s conceivable that \( P \). If you look at things just from the vantage point of classical propositional logic and the duality of diamond and box—regardless of how they are interpreted—(D) seems to be a weaker principle than (T) by way of entailing the latter without being entailed by it. From this perspective, the bare fact that realism is incompatible with (T) is not enough to show that realism is also incompatible with (D). However, on the intended interpretation, it’s easy to convince oneself that realism conflicts with (D) just as well. Only two quite modest principles are required to make that case. The first, already employed in the previous section, is (P1), stating the entailment from conceivability to thinkability. The second principle says that if it’s thinkable that not-\( P \), it’s also thinkable that \( P \):

\[
\text{(P2) } \Diamond \neg P \rightarrow \Diamond P
\]
Consider how mysterious it would be if the negation of $P$ was thinkable while $P$ itself for some reason wasn’t. (P2) plausibly rules out this constellation as inadmissible.

With these two principles in place, the incompatibility between (D) and realism can be shown as follows. Assume that, for some $P$, it’s unthinkable that $P$. By existential elimination, contraposition on (P2), and modus ponens it follows that $P$’s negation is also unthinkable: $\neg\Box\neg P$. By contraposition, (P1) entails $\neg\Box P \rightarrow \neg\Diamond P$. Instantiating this by substituting $\neg P$ for $P$, we get $\neg\Box\neg P \rightarrow \neg\Diamond\neg P$, which together with the above $\neg\Box\neg P$ entails that $P$’s negation is inconceivable as well, $\neg\Diamond\neg P$, or equivalently, $\Box P$. This verifies the antecedent of (D) for the considered $P$. But since $P$ is unthinkable, by the contrapositive of (P1) it follows that $P$ is also inconceivable, which falsifies the consequent of (D) for the considered $P$. In other words, if that which realism claims to be conceivable—viz. that there are (unthinkable truths and hence) unthinkable propositions—is true for some $P$, then that very $P$ will constitute a counterexample to (D): $\neg\Box P \rightarrow \neg(\Box P \rightarrow \Diamond P)$. Conversely, if (D) is a logical truth about conceivability, then we can apply modus tollens to reduce to absurdity the assumption that $P$ is unthinkable.

The situation is as before with (T). If the logic of conceivability is at least as strong as system K, then the realist’s conceivability claim (3) is logically inconsistent with (D). As noted, in Sect. 4 it will turn out that a certain species of realism rules against counting formula (K) as a logical truth, which might seem to make our K-based argument for the incompatibility between realism and (D) unacceptable to a proponent of that species of realism. But as in the case of (T), the incompatibility argument can be sustained by assuming instead of full (K) a restricted, realism-compatible version of (K); of which more in Sect. 5.

Above I said that from the perspective of propositional logic and duality, (D) seems to be a weaker principle than (T). However, given the following as a third background principle, (T) actually becomes derivable from (D):

\[(P \land \neg\Diamond P) \rightarrow \neg\Box P\]

Before saying what speaks for or against this principle, recall that given (D), it follows from the above reductio that $\Box P$, for arbitrary $P$. Taken together with (P3), this entails the negation of (P3)’s antecedent by modus tollens, $\neg(P \land \neg\Diamond P)$, which under the definition of $\rightarrow$ is equivalent to (T). Thus, in the presence of (P3), (T) and (D) turn out to be of equal strength.

But what reason is there to accept (P3)? Of course, inconceivability per se does not entail unthinkability. For instance, it’s inconceivable but not unthinkable that two plus two equals five. However, (P3) claims that the only way for a truth to be inconceivable is to also be unthinkable. And indeed, what reason could there be for a true proposition to be inconceivable other than its lack of thinkability? (Like (P1) and (P2), (P3) is neutral with regard to the realism/antirealism debate. Antirealists are committed to accepting (P3) as trivially true, since they maintain that its antecedent is false for all $P$, which makes the conditional vacuously true.)

That said, (P3) is bound to be less uncontroversial than (P1) and (P2). The most serious potential objection to (P3) stems from concern about the conceivability, or
lack thereof, of (so-far) undecidable non-contingent propositions, such as Goldbach’s Conjecture (GB). While one might be inclined to think that both GB and \( \neg \text{GB} \) are conceivable, Yablo (1993) argues that neither is conceivable in any sense involving “an appearance of possibility.” Roughly, the idea is that if one of GB and \( \neg \text{GB} \) is conceivable in that sense, so is the other; but then both should appear to be possible. This however would mean that GB appears to be contingent, which it most certainly does not. It then seems more plausible to say that neither GB nor \( \neg \text{GB} \) appears to be possible (which is not to say that they appear to be impossible), and accordingly that neither is conceivable. Now, given realism about mathematics, one of GB and \( \neg \text{GB} \) will be true, notwithstanding their undecidability. Let \( P \) stand for whichever of the two is true. Then, since both GB and \( \neg \text{GB} \) are thinkable, we have a counterexample to (P3).

Now, Yablo himself denies that conceivability \textit{qua} believability (of possibility) entails conceivability in the sense involving an appearance of possibility. If that is correct, the objection under consideration should not apply to (P3) with ✦ read as doxastic/imaginative conceivability, which are defined as believability and believability of possibility, respectively. But, as mentioned in Sect. 1, Yablo’s notion of believability is thoroughly negative, articulated in terms of the mere inability to rule out the proposition in question. One might reasonably worry that this construal, traditional as it may be, remains rather superficial. If there are unthinkable propositions, then, since we cannot rule out what we cannot even think,\(^9\) any such proposition would count as “believable,” negatively construed. If we rectify this defect by requiring that believability be understood in a sense that does involve an appearance of possibility, this exposes (P3) to Yablo’s objection. The same problem ensues if we alternatively add any other positive ingredient to the notion of believability that can be argued to fail to apply to both GC and its negation.

However, even then one might still wish to preserve what was plausible about (P3) by restricting it to contingent (and perhaps decidable non-contingent) propositions. The effect of this restriction would be that, over that domain, (D) and (T) turn out to be of equal strength, while (D) remains weaker than (T) in general, taking into account (undecidable) non-contingent propositions.

Be this as it may, we do not need (P3) to illustrate the incompatibility between realism and (D). To that end, the much less controversial (P1) and (P2) suffice. While (P3) will play no role in the following, the Yablo objection to its unrestricted version will resurface in Sect. 4 \textit{vis-à-vis} a kindred, fourth background principle.

### 4 The (K) formula

Last on our checklist is (K):

\[
\text{(K)} \quad \square (P \rightarrow Q) \rightarrow (\square P \rightarrow \square Q)
\]

\(^9\) Perhaps we can \textit{indirectly}—i.e. without thinking them—“rule out” \textit{inconsistent} unthinkable propositions, based on some general principle; but that still leaves all the consistent unthinkable propositions, if there are such.
On the intended interpretation, this says that if it’s inconceivable that \( P \) is the case without \( Q \) being the case, then if it’s inconceivable that \( P \) isn’t the case, it’s also inconceivable that \( Q \) isn’t the case. This rings plausible. And as far as I can tell, realism per se does not contradict that claim. However, if you combine realism with two further background principles, both showing some degree of intuitive plausibility, you end up with a view that is inconsistent with (K). The first is:

\[
(P4) \quad \Box P \rightarrow (\Diamond P \lor \Box \neg P)
\]

This says that if it’s thinkable that \( P \), then it’s either conceivable that \( P \) or conceivably that not-\( P \). While intuitively attractive, (P4) will be subject to the same Yablo-style objection as was (P3), discussed near the end of Sect. 3. And just as (P3), (P4) can be shielded against that objection by restricting it to contingent (and decidable necessary) propositions without interfering with present purposes.

The second additional principle is a generalization of (P2):

\[
(P2^*) \quad \Box \# P \rightarrow \Box P
\]

Here, \( \# \) schematically stands in for any unary open sentence of arbitrary finite complexity, i.e. any open sentence which, when saturated with one closed sentence, yields a closed sentence in turn. In the simplest case, \( \# \) will be negation, but it may also be something as contrived as ‘either snow is white or blank’, ‘if blank, then grass is green’ etc. (P2*) makes sense on a structural view of propositions. Structuralists hold that propositions have structure, roughly like sentences have structure. Although the analogy may not be perfect at the atomic level, the general idea is that, given e.g. a disjunctive proposition \( P \lor Q \), that proposition is a structured complex containing proposition \( P \) and proposition \( Q \) as sub-propositions. Likewise for \( P \rightarrow Q \) etc. (These examples allude to propositional-logical complexity, but similar examples could be given for predicate-logical complexity.) If propositions indeed are structured in this way, it’s quite natural to expect that a proposition is thinkable only if all of its sub-propositions are thinkable as well, and this is just what (P2*) hints at in its schematic way.

Given (P1), the structualist-minded (P2*), and (restricted) (P4), you can prove that, if that which realism claims to be conceivable is true for some \( P \), then that very \( P \), paired with any \( Q \) whose negation is conceivable, constitutes a counterexamples to (K). To simplify the proof, it will be expedient to begin by deriving two theorems. The first is equivalent to the converse of (P4), and follows from (P1) and the ungeneralized (P2):

\[
(Th1) \quad \neg \Box P \rightarrow (\neg \Diamond P \land \neg \Box \neg P)
\]

This states that unthinkability entails both inconceivability and inconceivability of the opposite. Proof: By (P2) and contraposition, \( \neg \Box P \rightarrow \neg \Diamond \neg P \). By (P1) and contraposition, \( \neg \Box P \rightarrow \neg \Diamond P \). Instantiating the latter with \( \neg P \) for \( P \) yields \( \neg \Box P \rightarrow \neg \Diamond \neg P \). By transitivity on the first and third of these conditionals, \( \neg \Box P \rightarrow \neg \Diamond \neg P \). Taken together with the second above conditional, this entails (Th1) by elementary propositional logic.
The second theorem is derivable from (P1), the generalized (P2*), and (restricted) (P4):

\[(\neg \square P \land \neg \neg \neg P) \rightarrow \neg \square \# P\]

This says that if both \(P\) and its negation are inconceivable, then any complex proposition of which \(P\) is a sub-proposition is also inconceivable. Proof: Take the contrapositive of (P4), \(\neg (\square P \lor \neg \square P) \rightarrow \neg \square P\), or equivalently \((\neg \square P \land \neg \square P) \rightarrow \neg \square P\). Assume the antecedent and apply modus ponens to get \(\neg \square P\). By (P2*) and contrasposition, \(\neg \square P \rightarrow \neg \square \# P\). Modus ponens then takes us to \(\neg \square \# P\). By (P1), this entails \(\neg \square \# P\). Eliminating our assumption results in (Th2). (If (P4) must be restricted, (Th2) will inherit that restriction.)

Let’s now put our two theorems to work to illustrate the discordance that ensues between realism and \((K)\) if the former is combined with (P2*) and (P4). Two assumptions are required. The first is that which realism claims to be conceivable, i.e. (1) \(\exists \square (P \land \neg \square P)\). By existential and conjunction elimination, \(\neg \square P\). By (Th1) and modus ponens, \(\neg \square P \land \neg \square \neg P\). By conjunction elimination and duality, \(\square P\). The second assumption is that some propositions are conceivably false:

\[(5) \quad \exists \square P \land \neg \square P\]

By existential elimination on (5), \(\neg \square Q\), or equivalently, \(\neg \square Q\). Conjoining this with the above \(\square P\), it follows that \(\square P \land \neg \square Q\). By the definition of \(\rightarrow\), \(\neg (\square P \rightarrow \square Q)\). Applying modus ponens to (Th2) and the above \(\neg \square P \land \neg \square \neg P\) yields \(\neg \square \# P\). Instantiating this by plugging \(\neg (\text{blank} \rightarrow Q)\) into \# gives us \(\neg \square (\neg P \rightarrow Q)\). By duality, \(\square (P \rightarrow Q)\). Conjoining this with the above \(\neg (\square P \rightarrow \square Q)\) results in \(\square (P \rightarrow Q) \land \neg (\square P \rightarrow \square Q)\). Applying the definition of \(\rightarrow\) once again yields \(\neg (\square [P \rightarrow Q] \rightarrow [\square P \rightarrow \square Q])\). Finally, existential introduction in effect turns this into the conclusion that there’s a counterexample to \((K)\).

With (P2*) and (P4) given, in order to avoid this conclusion one would have to reject one of its two assumptions. For the realist, to reject the first assumption would be to give up their position. The second assumption, (5), is hardly assailable, independent of one’s views about realism and antirealism. Clearly, it’s conceivable that Socrates isn’t wise, or that grass isn’t purple, etc. In order not to trivialize the notion of conceivability altogether, there must be such propositions. But if the realist accepts the proof’s assumptions as conceivable, and also accepts the background principles used to derive the failure of \((K)\) based on those assumptions, shouldn’t they also accept that it’s conceivable that \((K)\) fails, and hence that \((K)\) isn’t a logical truth? On the one hand, the answer is yes, if conceivability satisfies system \(K\). But on the other hand, to whatever extent one considers the proof a refutation of \((K)\), one cannot accept that conceivability satisfies \(K\), since the latter treats \((K)\) as an axiom. However, as we shall see in a moment, the full strength of \((K)\) is not required to expand the proof into an argument for the conceivable failure of \((K)\).
5 Realist restrictions

Each of the considered arguments against (T), (D), and (K) depended essentially on the claim that it’s conceivable that there are unthinkable propositions, a direct consequence of realist claim (3). Accordingly, as far as they go, those arguments don’t speak against the results of restricting (T), (D), and (K) by a thinkability condition:

\[(T) \quad \Box P \rightarrow (\Box P \rightarrow P)\]
\[(D) \quad \Box P \rightarrow (\Box P \rightarrow \Diamond P)\]
\[(K) \quad (\Box P \land \Box Q) \rightarrow (\Box (P \rightarrow Q) \rightarrow [\Box P \rightarrow \Box Q])\]

In Sects. 2 and 3 we noted that realism is inconsistent with unrestricted (T) and (D), provided i.a. that the dual of conceivability satisfies unrestricted (K). In Sect. 4 it turned out that this condition is in tension with at least one version of realism. But even proponents of that version are in a position to agree that, instead of taking their objection against (K) to be grounds for simply rejecting the latter without compensation, it seems more commensurate to restrict (K) in a way that rescues what was initially plausible about it. The case is analogous to the restriction of necessitation in the face of two-dimensional counterexamples, touched upon in Sect. 2. In the present case, the obvious restriction would be (K\textsubscript{\textbullet}). It’s easy to see that the latter is still strong enough to sustain the incompatibility arguments from Sects. 2 and 3. Together with \Box necessitation, (K\textsubscript{\textbullet}) defines what may be called system K\textsubscript{\textbullet}, which allows us to infer \Diamond Q from \Diamond P whenever (a) Q is derivable from P and (b) \Box P and \Box Q are given. Since (1) \exists P(P \land \neg \Box P) entails (2) \exists P(P \land \neg \Diamond P), and since obviously enough both (1) and (2) are thinkable, (3) \Diamond \exists P(P \land \neg \Diamond P) may be used as a premise to derive (4) \Diamond \exists P(P \land \neg \Diamond P). Since the realist is committed to (3), and since (3)’s consequence (4) contradicts unrestricted (T)/(T\textsuperscript{\textcircled{}}}), realism is incompatible with the latter. The argument for the incompatibility between realism and unrestricted (D) from Sect. 3 can be modified in an analogous way, again without using full (K), and thus without begging the question against the realist argument from Sect. 4.

Finally, to turn the latter into a more direct argument for the incompatibility between (K) and the version of realism that accepts (P2\textsuperscript{*}) and (P4), recall the two assumptions the argument started with, (1) \exists P(P \land \neg \Box P) and (5) \exists P\neg \Box P, as well as the conclusion it led to, \exists P\exists Q\neg ([\Box P \rightarrow Q] \rightarrow [\Box P \rightarrow \Box Q]). Since, what may be derived from a set of premises may also be derived from their conjunction, let’s recast the two assumptions as one conjunctive assumption. A conjunction whose conjuncts are conceivable is conceivable itself (setting aside potential worries about complexity aggregation, which don’t apply in our case). Since the realist considers the (1) conjunct conceivable, and since the (5) conjunct should be conceivable by anybody’s lights, the realist who accepts (P2\textsuperscript{*}) and (P4) should be willing to concede the conceivability of the conjunction, \Diamond (\exists P[P \land \neg \Box P] \land \exists P\neg \Diamond P). Since both the conjunctive assumption as well as the anti-(K) conclusion are thinkable, in K\textsubscript{\textbullet} that concession commits one to also accepting \Diamond \exists P\exists Q\neg ([\Box P \rightarrow Q] \rightarrow [\Box P \rightarrow \Box Q]), i.e. that (K) is conceivably false, and hence not a logical truth.
We have been concerned with (T), (D), and (K) as implicitly quantified claims, not as sentence schemata where ‘P’ and ‘Q’ are placeholders for arbitrary sentences of some language. It’s worth pointing out that the considered realist arguments against those claims do not speak against the suggestion that, understood as schemata, (T), (D), and (K) are true under any uniform substitution of sentences for schematic sentential variables. Call a schema that satisfies this condition substitution valid. Assume for a moment that the unthinkability considerations offered by the realist are the only reason for the failure of quantified (T), (D), and (K). Then producing a false substitution instance of, say, the (T) schema would require plugging a sentence into its ‘P’ positions that expresses an unthinkable proposition; likewise for schematic (D) and (K). However, no such sentence exists, assuming secondly that expressibility entails thinkability. (Set aside potential worries about complexity-related limits on thinkability, to the effect that some propositions are too complex to be thinkable, while still being expressible.) Given both assumptions, it follows that the (T) schema is substitution valid; and likewise for schematic (D) and (K). But even if realism is compatible with thinking that all substitution instances of these schemata are true, it does disincentivize considering said substitution instances logically true. After all, there is something to the idea that a sentence expresses a logical truth in the purest sense only if its universal generalization does so as well; as witnessed by the circumstance that in most systems of nth-order quantified (modal) logic a schema is valid/derivable only if the corresponding nth-order universal generalization is valid/derivable as well.

In standard model theories for modal logics, (T) and (D) encode potential properties of the accessibility relation between evaluation points. In particular, (T) corresponds to the property of reflexivity, while (D) corresponds to seriality. It follows that no standard model theory for conceivability that makes accessibility reflexive and/or serial is acceptable to the realist.10 In the relevant sense, a model theory is standard provided it treats a formula governed by diamond (box) as true at a point iff the formula immediately embedded under that operator is true at some (all) accessible points. Instead of selecting a non-reflexive and non-serial accessibility relation, one idea to make the model theory for ♦ and □ realistically acceptable would be to enrich ordinary models <W, R, I> with a set of (structured) propositions, together with a subset of the former, T, comprising the thinkable propositions. What this would allow one to do is to supplement the standard clause for ♦ with the additional requirement that the proposition expressed by the formula embedded under ♦ be an element of T. (The obvious clause for □ would be that \( \Box P \) is true at w iff \( P \in T \).) That said, if the model theory also contains standard clauses for alethic \( \Diamond \) and \( \Box \), then, while that idea would be enough to invalidate the antirealist inference from possibility to conceivability, further measures would have to be taken in order not to validate the equally controversial inference from conceivability to possibility. Since the formal semantics for conceivability is not the topic of this paper, I’ll leave that discussion for another

10 In contrast, realism does not rule out either of the (IV) formula: ♦♦P → ♦P, the (B) formula: \( P \rightarrow \Box \Diamond P \), or the (V) formula: ♦P → ♦♦P, which correspond, respectively, to transitivity, symmetry, and the Euclidean property.
occasion. What may be safely asserted at this point is that, in the presence of (P1), (P2*), and (P4), any logic of conceivability that would be acceptable by realist standards is bound to be non-normal, due to the failure of (K).

6 Concluding remarks on apriority and inconceivability of the opposite

We set up the conceivability interpretation of diamond and box by starting with the notion of conceivability, assigning that notion to diamond, and defining box as its dual. We then made the following observations about the realist’s claim that it’s conceivable that there are unthinkable truths. First, the claim is inconsistent with (T), given (K), □ necessitation, and (P1). Second, given additionally (P2), the claim is inconsistent with (D). Third, given also generalized (P2*) and (restricted) (P4), the claim is inconsistent with (K).

Note that if (T) and (D) fail on the conceivability interpretation, the dual of conceivability, i.e. inconceivability of the opposite, is extremely weak. Not only does it fail to entail truth, due to the failure of (T); it does not even entail conceivability, due to the failure of (D). But then there is a straightforward argument to be made that inconceivability of the opposite fails to entail, let alone be tantamount to, apriority; at least if we understand ‘it’s a priori that \( P \)' in terms of ‘it’s knowable a priori that \( P \)', as seems to be standard, and read ‘knowable’ in its factive sense. (See footnote 5.) On this reading, apriority is factive due to the factivity of knowability, and also obviously entails conceivability since whatever is knowable is also conceivable. As a consequence, if (T) and (D) fail, apriority is not the dual of conceivability, nor is it entailed by that dual.

An alternative approach to the logic of broadly epistemic modality would start with the notion of apriority, assign it to box, and define diamond as its dual: ‘it’s not a priori that not-\( P \)’. Given that starting point, (T), (D), and (K) are much less contestable. At any rate, nothing extractable from specifically realist reasoning can be used against them on this apriority interpretation. Given that ‘it’s a priori that \( P \)' is understood as above, (T) and (D) are validated at once. However, now it’s the dual of apriority that becomes excessively weak by the realist’s lights, in virtue of not even entailing, let alone being tantamount to, conceivability. If there are unthinkable propositions, as the realist thinks is conceivable, then any such proposition will trivially satisfy the dual of apriority. After all, if it’s not even thinkable that \( P \), then not-\( P \) will be unthinkable as well, hence unknowable, hence unknowable a priori.

This has consequences for the conventional approach to explicate (one notion of) conceivability by defining ‘it’s conceivable that \( P \)’ as ‘it’s not a priori that not-\( P \)’. If realism is correct, conceivability so defined fails to entail thinkability, by the above. To be sure, those working with that definition, such as Chalmers (2002), routinely qualify it by calling the definiendum a “negative” notion of conceivability (contrasting it with a notion on which the conceivability of \( P \) requires the feasibility of forming some “positive conception” of a situation in which \( P \)). But Chalmers presumably does not wish his definiendum to be quite as negative as it turns out to be, considered from a realist vantage point. Arguably, a notion of conceivability on
which the latter does not even require thinkability is a notion of conceivability in name only. Anyway, to suppose that conceivability, in any interesting sense, is definable along the lines of the above recipe, is to define one’s way towards an all too easy argument against realism.

It may be urged that realists owe us an explanation as to why the explication of conceivability as the dual of apriority has been so attractive to philosophers. But it seems easy for the realist to explain away the intuition underlying that explication: Apriority and (doxastic) inconceivability of the opposite coincide over the domain of thinkable propositions. The intuition in question rests on the circumstance that we normally are not concerned with the speculative possibility of there also being unthinkable propositions, and it is the latter with respect to which the extensions of the two notions would come apart. This explanation seems sufficient to dismantle the intuition as a threat to realism.

Acknowledgements
Versions of this paper were presented at the workshop Logics for Imagination, Ruhr-University Bochum (March 2018), at the University of Konstanz (May 2018), at the Massachusetts Institute of Technology (May 2019), and at the University of California, Irvine (June 2019). My thanks to all four audiences, in particular to Christopher Badura, Lisa Benossi, Francesco Berto, Daniel Dohrn, Sophia Gilbert, Caspar Hare, Vann McGee, Toby Meadows, Thomas Müller, Eric Raidl, Pierre Saint-Germier, Mattias Skipper, Jack Spencer, Will Stafford, Kai Wehmeier, and Roger White. Special thanks to Ali Esmi, Stephen Yablo, and an anonymous referee for Philosophical Studies.

Compliance with ethical standards
Conflict of interest The author declares that they have no conflict of interest.

References
Brogaard, B., & Salerno, J. (2008). Knowability, possibility and paradox. In V. F. Hendricks & D. Pritchard (Eds.), New waves in epistemology (pp. 270–299). Basingstoke: Palgrave Macmillan.
Casullo, A. (2003). A priori justification. Oxford: Oxford University Press.
Chalmers, D. J. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), Conceivability and possibility (pp. 145–200). Oxford: Oxford University Press.
Dummett, M. (1976). Truth and meaning. Oxford: Clarendon Press.
Evans, G. (1979). Reference and contingency. The Monist, 62(2), 161–189.
Fuhrmann, A. (2014). Knowability as potential knowledge. Synthese, 191(7), 1627–1648.
Gendler, T. S., & Hawthorne, J. (Eds.). (2002). Conceivability and possibility. Oxford: Oxford University Press.
Hofweber, T. (2019). Idealism and the Harmony of thought and reality. Mind, 128(511), 699–734.
Kripke, S. A. (1980). Naming and necessity. Cambridge, MA: Harvard University Press.
James, W. (1909). The meaning of truth. New York: Longmans, Green and Co.
McDowell, J. H. (1994). Mind and world. Cambridge, MA: Harvard University Press.
Nagel, T. (1986). The view from nowhere. New York: Oxford University Press.
Peirce, C. S. (1878). How to make our ideas clear. Popular Science Monthly, 12(Jan), 286–302.
Putnam, H. (1981). Reason, truth, and history. Cambridge, MA: Cambridge University Press.
Sorensen, R. A. (1988). Blindspots. Oxford: Clarendon Press.
Strawson, P. F. (1966). The bounds of sense: An essay on Kanti’s Critique of pure reason. London: Methuen.
Williamson, T. (2000). Knowledge and its limits. Oxford: Oxford University Press.
—— (2007). The philosophy of philosophy. Malden, MA: Blackwell Publishing.
Wittgenstein, L. (1975). Philosophical remarks. Oxford: Basil Blackwell.

Springer
Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, 53(1), 1–42.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.