doi: 10.5007/1808-1711.2017v21n1p41

IDEAL REASONERS DON'T BELIEVE IN ZOMBIES

DANILO FRAGA DANTAS

Abstract. The negative zombie argument concludes that physicalism is false from the premises that $p \wedge \neg q$ is ideally negatively conceivable and that what is ideally negatively conceivable is possible, where p is the conjunction of the fundamental physical truths and laws and q is a phenomenal truth (Chalmers 2002; 2010). A sentence ϕ is ideally negatively conceivable iff ϕ is not ruled out a priori on ideal rational reflection. In this paper, I argue that the negative zombie argument is neither a priori nor conclusive. First, I argue that the premises of the argument are true only if there exists an adequate finite ideal reasoner $\mathscr{R}$ that believes $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$ on a priori basis. Roughly, a finite reasoner is a reasoner with cognitive limitations (e.g. finite memory). I argue that $\mathscr{R}$ is finite only if $\mathscr{R}$ reasons nonmonotonically and only approach ideal reflection at the limit of a reasoning sequence. This would render the argument nonconclusive. Finally, I argue that, for some q, $\mathscr{R}$ does not believe $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$ on a priori basis (e.g. for $q =$'something is conscious'). This would render the choice of an adequate q dependent on empirical information (and the argument a posteriori). I conclude that the negative zombie argument (and, maybe, all zombie arguments) is neither a priori nor conclusive.

Keywords: Modal epistemology; zombie argument; finite reasoning; nonmonotonic reasoning.

Introduction

The conceivability principle (cp, 'if ϕ is conceivable, then ϕ is possible') is often regarded as basing nontrivial modal knowledge that $\Diamond\phi$ (see Gendler and Hawthorne 2002).[1] In philosophy of mind, for example, this principle is used in the *zombie argument* (Chalmers 2002; 2010):

(p1)	$p \wedge \neg q$ is conceivable.
(p2)	If ϕ is conceivable, then ϕ is possible. (cp)
(p3)	If $p \wedge \neg q$ is possible, then physicalism is false.
$\therefore$	Physicalism is false.

In the zombie argument, p is the conjunction of the fundamental physical truths and laws and q is an arbitrary phenomenal truth (Chalmers 2010, p.107).[2] If q expresses the truth that something is conscious, then $\Diamond(p \wedge \neg q)$ asserts the possibility of a zombie world, i.e. a world physically identical to the actual world, but in which nothing is conscious.

Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

There exists an extensive literature about the zombie argument (see Kirk 2012). Premise p3 is relatively undisputed: it is often accepted that physicalism entails that the phenomenal truths supervene on the physical truths (McLaughlin and Bennett 2014, sec.5.2).[3] Premises p1 and p2 are disputed. Chalmers defends p1 and p2 by distinguishing different senses of 'conceivable' and 'possible' and by pointing senses in which $p \wedge \neg q$ would be conceivable and conceivability would entail possibility. Among other things, Chalmers defends the *negative zombie argument* (Chalmers 2010, p.144):

(p1) $p \wedge \neg q$ is ideally negatively conceivable.
(p2) If ϕ is ideally negatively conceivable, then ϕ is possible. (cp)
(p3) If $p \wedge \neg q$ is possible, then physicalism is false.
$\therefore$ Physicalism is false.

In the following, p1 and p2 refer to premises p1 and p2 of the *negative* zombie argument. Accordingly, cp refers to its form in p2: 'if ϕ is *ideally negatively* conceivable, then ϕ is possible'.

The notion of ideal negative conceivability used in this argument is explained as follows:

Definition 1 (Ideal negative conceivability 1) *A sentence ϕ is ideally negatively conceivable iff ϕ is not ruled out a priori on ideal rational reflection (Chalmers 2002, pp.143–4).*

In this paper, I argue that the negative zombie argument is neither a priori nor conclusive. In section 1, I argue that premises p1 and p2 are both true only if there exists a finite ideal reasoner with the relevant properties that believes $\Diamond(p \wedge \neg q)$ on the basis of not ruling out $p \wedge \neg q$ on a priori basis. Roughly, a finite reasoner is a reasoner with cognitive limitations (e.g. finite memory). In section 2, I argue that an ideal reasoner with the relevant properties cannot be finite, what would undermine the use of cp in the negative zombie argument. In section 3, I argue that there exists a related notion of ideal negative conceivability for which there exists a finite nonmonotonic ideal reasoner with the relevant properties, but that this reasoner negatively conceiving $p \wedge \neg q$ is only nonconclusive reason for $\Diamond(p \wedge \neg q)$. In this case, the negative zombie argument would not be conclusive. In section 4, I argue that this reasoner does not negatively conceive $p \wedge \neg q$ for every choice of q. For example, this reasoner does not negatively conceive $p \wedge \neg q$ for $q =$'something is conscious'. The choice of an adequate q would depend on empirical information and this would render the argument a posteriori. I conclude that the negative zombie argument (and, maybe, all zombie arguments) is neither a priori nor conclusive.

1. Ideal negative conceivability

The major difficulty in evaluating the negative zombie argument is the notion of ideal negative conceivability used in p1 and p2 not being clear. More specifically, the meaning of the phrase 'a priori ideal rational reflection' is not clear. This problem may be avoided by using a more precise notion of an ideal reasoner as a model of ideal rational reflection. This reading is implicit in Chalmers (2010), but explicit in Menzies (1998):

> Under what circumstances do our corrective practices discount acts of conceiving as not being veridical indicators of possibility? The answer is simple: When they suffer from one kind of cognitive limitation or other. Let us call a subject who does not suffer any of those limitations an ideal conceiver. These reflections suggest a biconditional of the following kind for the concept of possibility: it is possible that ϕ iff an ideal conceiver could conceive that ϕ (Menzies 1998, pp.268–9).

The notion of an ideal conceiver explicit in Menzies (1998) and implicit Chalmers (2010, p.143) is (simply?) of a reasoner without cognitive limitations (e.g. unlimited memory). This notion cannot be used as a model of ideal rational reflection unless it specifies how that reasoner reasons. In the literature, there exists a notion of an ideal reasoner as a reasoner that believes all the logical consequences of its epistemic situation and has a nontrivial set of beliefs, where the epistemic situation of a reasoner is the information available for reasoning (e.g. explicit beliefs).[4] This notion is also not adequate for modeling ideal rational reflection because an ideal reasoner with these features may have random beliefs (depending on the notion of beliefs used, see sec. 3). The notion of a *strictly* ideal reasoner does not present this problem and may be used as a model of ideal rational reflection:

Definition 2 (Strictly ideal reasoner) *A reasoner $\mathscr{R}$ is strictly ideal iff:*

(i1) $\mathscr{R}$ believes all and only the logical consequences of its epistemic situation;

(i2) $\mathscr{R}$ has a nontrivial set of beliefs.

The notion of a strictly ideal reasoner is relative not only to a notion of beliefs but also to a logic because the notions of logical consequence and triviality are relative to a logic. For this reason, I will always talk about a strictly ideal reasoner for $\models^x$, where $\models^x$ is the *semantic* consequence relation of a logic x[5]. An a priori reasoner has the empty set as (initial) epistemic situation. In the following, 'ideal reasoner' refers to an a priori strictly ideal reasoner.

In this context, I argue that p1 and p2 are both true only if Z is true (i.e. p1 $\wedge$ p2 $\rightarrow$ Z):

(Z) There exists a finite ideal reasoner for an adequate $\models^x$ that believes $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$ (see fn. 6),

Sentence $p \wedge \neg q$ is ideally negatively conceivable iff $p \wedge \neg q$ is not ruled out on a priori ideal rational reflection (def. 1). I am supposing that an ideal reasoner for an adequate $\models^x$ is an adequate model of a priori ideal rational reflection. Then $p \wedge \neg q$ is ideally negatively conceivable iff $p \wedge \neg q$ is not ruled out by an ideal reasoner for $\models^x$. A sentence ϕ is ruled out by an ideal reasoner for $\models^x$ iff that reasoner believes $\neg\phi$.[6] Then (p1) $p \wedge \neg q$ is ideally negatively conceivable iff an ideal reasoner for $\models^x$ does not believe $p \rightarrow q$, which is equivalent to $\neg(p \wedge \neg q)$. Finally, (p2) ideal negative conceivability entails possibility for the case of $p \wedge \neg q$ only if an ideal reasoner for $\models^x$ believes $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$.

The claim that p1 and p2 are both true only if Z is true depends on the properties of $\models^x$. As I see it, Chalmers is correct when he claims that the adequate logic for evaluating the negative zombie argument is *the* logic of a priori (modal) knowledge.[7] According to Chalmers, ideal negative conceivability is a general principle of modal epistemology for generating nontrivial modal knowledge. In this context, an adequate $\models^x$ must fulfill the following requirements:

(r1) $\Diamond(p \wedge \neg q)$ is expressible in the language of $\models^x$;

(r2) For some ϕ, $\models^x \Diamond\phi$ and $\not\models^x \phi$;

(r3) For all ϕ, if $\not\models^x \neg\phi$, then $\models^x \Diamond\phi$;

(r4) $\models^x$ tracks the *metaphysical* modalities correctly;

(r5) For all *superphysical* truths ϕ, $\models^x p \rightarrow \phi$.[8]

Requirement r1 assures that an ideal reasoner for $\models^x$ has the required concepts for its use in the negative zombie argument (the concepts of possibility and necessity, the concepts used in p, the concepts used in q, etc). If an ideal reasoner for $\models^x$ is supposed to believe $\Diamond(p \wedge \neg q)$, then it must be able to believe $\Diamond(p \wedge \neg q)$. An ideal reasoner for $\models^x$ is able to believe $\Diamond(p \wedge \neg q)$ iff $\Diamond(p \wedge \neg q)$ expressible in the language of $\models^x$ (def. 2, i1). Then $\Diamond(p \wedge \neg q)$ must be expressible in the language of $\models^x$. Then r1 has the consequence that the language of $\models^x$ must contain modal operators and quantifiers.[9] For simplicity, I will suppose that $\models^x$ must be a modal extension of classical logic, meaning first-order logic (supraclassicality, but see fn. 15).

Requirement r2 assures that an ideal reasoner for $\models^x$ is able to have nontrivial modal knowledge. An ideal reasoner has nontrivial modal knowledge that $\Diamond\phi$ only if it believes $\Diamond\phi$ and does not believe ϕ (see fn. 1, knowledge implies belief). An ideal reasoner for $\models^x$ is able to, for some ϕ, believe $\Diamond\phi$ while not believing ϕ iff, for some ϕ, $\models^x \Diamond\phi$ and $\not\models^x \phi$ (def. 2, i1). Requirement r3 assures that an ideal reasoner for $\models^x$ employs negative conceivability as a *general* principle in generating

modal beliefs. An ideal reasoner employs negative conceivability for generating the belief that $\Diamond\phi$ iff it believes $\Diamond\phi$ on the basis of not believing $\neg\phi$ (see fn. 6). An ideal reasoner for $\models^x$ is able to employ negative conceivability as a general principle for generating modal beliefs iff, for all ϕ, if $\not\models^x \neg\phi$, then $\models^x \Diamond\phi$ (def. 2, i1). Together, r2 and r3 assure that an ideal reasoner for $\models^x$ is able to employ negative conceivability as a general principle for generating nontrivial modal beliefs (and, possibly, nontrivial modal knowledge).

Requirement r4 assures that an ideal reasoner for $\models^x$ knows how to reason about metaphysical modalities. An ideal reasoner for $\models^x$ knows how to reason about metaphysical modalities iff $\models^x$ tracks the metaphysical modalities correctly (def. 2, i1). In the literature, it is usually accepted that the metaphysical modalities have two *minimal* features (see Salmon 1989). First, all logical truths are necessary. Then $\models^x$ must be such that, for all ϕ, if $\models^x \phi$, then $\models^x \Box\phi$ (necessitation). Second, what is actual is possible (or, equivalently, what is necessary is actual). Then $\models^x$ must be such that, for all ϕ, $\models^x \phi \rightarrow \Diamond\phi$ or, equivalently, $\models^x \Box\phi \rightarrow \phi$ (reflexivity).

Requirement r5 assures that an ideal reasoner for $\models^x$ has the inferential capacities required for its use in the negative zombie argument. The negative zombie argument is such that if $p \wedge \neg q$ is ideally negatively conceivable, then q is not superphysical. The argument is sound only if, for all superphysical truths ϕ, $p \wedge \neg\phi$ is not ideally negatively conceivable (contraposition). Then an ideal reasoner for $\models^x$ must believe $p \rightarrow \phi$ for all superphysical truths ϕ. An ideal reasoner for $\models^x$ believes $p \rightarrow \phi$ for all superphysical truths ϕ iff, for all superphysical truths ϕ, $\models^x p \rightarrow \phi$ (def. 2, i1). Then $\models^x$ must be such that, for all superphysical truths ϕ, $\models^x p \rightarrow \phi$.

2. Conclusive ideal conceiver

The first step in evaluating Z is to investigate the features of a $\models^x$ that fulfills requirements r1-r5. There are modal extensions of classical logic that fulfill r1-r4, but not a $\models^k$, a Kripkean (normal) modal extension of classical logic (see Kripke 1959). Independently of the accessibility relation used, a $\models^k$ does not fulfill r1-r4 (specially r2 and r3) because $\models^k \Diamond\phi$ is false in all cases. Let $\models^r$ be a Carnapian modal extension of classical logic (see Carnap 1946). The difference between $\models^r$ and a Kripkean (normal) semantics is that $\models^r$ has a fixed frame $\langle W, R\rangle$ with a fixed model $\langle W, R, \Vdash\rangle$, where the possible worlds $w \in W$ represent all possible (consistent) assignments for all the atomic sentences in the language and R contains all possible pairs of worlds $w, w' \in W$ (fully connected). The interpretation of the logical connectives and modal operators in $\models^r$ are the same as in $\models^k$. Requirements r1-r4 are fulfilled by $\models^r$. First, $\models^r$ is a consistent modal extension of classical logic.[10] Then $\models^r$ fulfills r1. Second, for all atomic ϕ, $\models^r \Diamond\phi$ and $\not\models^r \phi$.[11] Then $\models^r$ fulfills r2. Third, for all ϕ, if $\not\models^r \neg\phi$, then

$\models^r \Diamond\phi$.[12] Then $\models^r$ fulfills r3. Finally, $\models^r$ fulfills both necessitation and reflexivity.[13] Then $\models^r$ fulfills r4.

However, an ideal reasoner for $\models^r$ cannot be finite. More generally, an ideal reasoner for a $\models^x$ that fulfills r1-r4 cannot be finite (Cohnitz 2012, p.70). A finite reasoner is a reasoner with finite cognitive capacities, such as finite memory and being able to execute only a finite number of inferential steps in a finite time interval (finite inferential capacities, see def. 6). That a finite ideal reasoner has finite inferential capacities may be modeled by requiring $\models^x$ to be recursively enumerable (Boolos et. al. 2007, p.3).[14] (i) A $\models^x$ that fulfills r1-r4 is such that, for all ϕ, $\models^x \phi$ iff $\models^x \Box\phi$. If $\models^x \phi$, then $\models^x \Box\phi$ (r4, necessitation). Suppose that $\models^x \Box\phi$. Then $\models^x \phi$ (r4, reflexivity). Therefore, $\models^x \phi$ iff $\models^x \Box\phi$. Also, (ii) a $\models^x$ that fulfills r1-r4 is such that, for all ϕ, $\not\models^x \phi$ iff $\models^x \Diamond\neg\phi$. If $\not\models^x \phi$, then $\models^x \Diamond\neg\phi$ (r3, $\phi/\neg\phi$). Suppose that $\models^x \Diamond\neg\phi$. Then $\models^x \neg\Box\phi$ (definition of $\Diamond$). Then $\not\models^x \Box\phi$ (r1, consistency).[15] Then $\not\models^x \phi$ (i). Therefore, $\not\models^x \phi$ iff $\models^x \Diamond\neg\phi$. Suppose that $\mathscr{R}^x$ is an ideal reasoner for a $\models^x$ that fulfills r1-r4. Now, suppose that $\mathscr{R}^x$ is finite. Then $\models^x$ is recursively enumerable (finite reasoner). For infinitely many ϕ, $\models^x \Diamond\neg\phi$ (r2, $\phi/\neg\phi$, see fn.)[16] and $\models^x \Diamond\neg\phi$ iff $\not\models^x \phi$ (ii). Then $\models^x$ is recursively enumerable iff $\models^x$ is recursively decidable.[17] If $\models^x$ is decidable, then classical logic is decidable (r1, supraclassicality). But classical logic is not decidable (Church 1936). Therefore, $\mathscr{R}^x$ is not finite.

There exists a different reason for which an ideal reasoner for a $\models^x$ that fulfills r1-r5 is not finite. Requirement r5 states that $\models^x$ must be such that, for all superphysical truths ϕ, $\models^x p \rightarrow \phi$, where p is the conjunction of the fundamental physical truths and laws. Quantum mechanics (QM) is usually thought to be the best candidate for the fundamental physical description of the world. But QM is usually thought to have an indeterministic character in the sense that the laws of QM and the full description of a quantum system at t_1 do not determine the full state of the system at $t_2 > t_1$.[18] In this context, the relation between a p with finitely many conjuncts and the complete set of physical truths (which are superphysical) cannot be deductive[19]. For this reason, a $\models^x$ which fulfills requirement r5 must be itself nondeductive (e.g. nonmonotonic). However, an ideal reasoner for a well-behaved nonmonotonic $\models^x$ cannot be finite.[20]

In considering this argument, Chalmers replies that the conclusion that $\models^x$ must be nondeductive may be avoided by adding to p what he calls 'interpretative principles':

> On the collapse interpretation [of QM], a natural interpretative strategy is to say that an entity is located in a certain region of three-dimensional space if a high enough proportion of the (squared) amplitude of its wave-function is concentrated within that region. ... An interpretative principle involving 'a high enough proportion' will then deliver classical truth at both the microscopic and macroscopic level (Chalmers 2012, pp.294–5).

But Chalmers's interpretative principles are themselves nondeductive. In the example, from the fact that there exists a region where a high enough proportion of the (squared) amplitude of a wave-function is concentrated, it does not follow (deductively) that an entity is located in that region (argument due to Parent 2016, p.239). Then a $\models^x$ that verifies such interpretative principles must be nondeductive (e.g. nonmonotonic) and the conclusion is not avoided.

Chalmers also argues that the conclusion may be avoided by adding more information to p:

> The apparent failure of determinism in quantum mechanics suggests that the [Laplace's] demon [the ideal reasoner] could not predict the future just from facts about physical laws and about the present. . . . To avoid these problems, however, we need only give Laplace's demon more information than Laplace allows. To accommodate nondeterminism, we might give the demon full information about the distribution of the fundamental physical entities throughout space and time (Chalmers 2012, p.xiv).

But, if p contains "the full information about the distribution of the fundamental physical entities through space and time", then p must have infinitely many conjuncts. This is the case because time is often modeled as dense (see Ray 1991, p.20) and p would need to contain information about the position of the fundamental physical entities for infinitely many moments between every t_1 and t_2. In this case, an ideal reasoner for $\models^x$ would not be finite as well. A finite reasoner has finite memory, but p would be an infinitary conjunction. This means that a finite reasoner would not be able to entertain the sentences p or $p \rightarrow \phi$, much less reason about them (see def. 6). Then requirement r5 has the consequence that either $\models^x$ is nondeductive (e.g. nonmonotonic) or p is infinitary. In both cases, an ideal reasoner for a $\models^x$ cannot be finite.

This conclusion seems to be accepted by Chalmers (2010) and Menzies (1998), which presuppose an ideal conceiver without cognitive limitations in their notions of ideal (negative) conceivability (see sec. 1). Chalmers (2010), specifically, does not seem to find this conclusion problematic. I find two issues with the conclusion that the ideal conceiving necessary for the negative zombie argument cannot be performed by a finite reasoner. The first issue is: how finite reasoners, such as humans, could have a clear grasp of whether $p \wedge \neg q$ is *ideally* conceivable? If this grasp is somehow 'intuitive' (as Chalmers 2010, p.155 suggests), it is not clear which is the role of cp in the argument. The second (and more important) issue is that cp is supposed to have a normative role in modal epistemology (i.e. basing nontrivial modal knowledge), but the (rational) patterns of inference for a reasoner without cognitive limitations may be fundamentally different from those of a finite reasoner (e.g. humans). For example, a reasoner which has full information about the distribution of

the fundamental physical entities through space and time 'predicts' the position of a fundamental physical entity at t simply by reinstating what it already knows and without making use of the physical laws, what is fundamentally different from how finite reasoners reason about physics. For another example, if there exists a procedure for checking guesses, a reasoner which is able to perform inferences instantaneously may solve any problem (instantaneously) simply by generating and checking random guesses. If cp has a normative role in modal epistemology, then the patterns of inference of an ideal conceiver should serve as parameter of rationality for, for example, humans (finite reasoners). But an ideal conceiver that is adequate for the negative zombie argument cannot be finite and the example show that the (rational) patterns of inference of a reasoner without cognitive limitations may hardly be seem as a parameter of rationality for finite reasoners.

3. Nonconclusive ideal conceiver

Occasionally, Chalmers proposes a slightly different notion of ideal negative conceivability:

Definition 3 (Ideal negative conceivability 2) *A sentence ϕ is ideally negatively conceivable iff ϕ is negatively conceivable with justification that is undefeatable by better reasoning (Chalmers 2002, p.172),*

where ϕ is negatively conceivable when ϕ is not ruled out on a priori basis.

The notion of undefeatability by better (further) reasoning is often used in the literature on nonmonotonic reasoning.[21] In the theory of defeasible (nonconclusive) reasoning (Pollock 1990; 1995), for example, there exists a notion of warrant which is closely related to undefeatability by better reasoning. The theory describes reasoning as the procedure of adopting beliefs on the basis of reasons and retracting beliefs on the basis of defeaters. Reasons are sets of mental states that provide epistemic support for beliefs. A reason may be conclusive or nonconclusive depending on the epistemic support that it provides. The epistemic support of conclusive reasons cannot be overridden by new information (defeaters). For example, believing ϕ is a conclusive reason for believing $\phi \vee \psi$. The epistemic support of nonconclusive reasons can be overridden by defeaters. For example, perceiving an object as red is nonconclusive reason for believing that the object is red. The belief that an object is red may be defeated by the information that the object is under red light.[22]

Let the $\mathtt{KB}_i$ be sets of reasons and $\mathtt{KB}_0, \mathtt{KB}_1, \ldots, \mathtt{KB}_i, \ldots$ be the reasoning sequence of a reasoner (how the reasoner would reason from the available information if it had enough cognitive resources, e.g. memory). Then the notion of warrant is the following:

Definition 4 (Warrant) *A belief ϕ is warranted relatively to $\mathtt{KB}_0, \mathtt{KB}_1, \ldots, \mathtt{KB}_i, \ldots$ iff there exists an i such that, for all $j \geq i$, ϕ is undefeated relatively to $\mathtt{KB}_j$ (Pollock 1995, p.133).*

In other words, the belief that ϕ is warranted iff there exists a stage in the reasoning sequence such that the belief is undefeated at every subsequent stage. Then the belief that ϕ is undefeatable by better (further) reasoning iff that belief is warranted.

Consider a model of an a priori reasoner that provides a clear definition for the notion of a reasoning sequence and the different notions of beliefs. In the model, a reasoner is composed of a language ($\mathscr{L}$), a knowledge base (KB), and a pattern of inference (π), where KB is a set of sentences in $\mathscr{L}$ that models the (initial) epistemic situation of the reasoner and $\pi : 2^{\mathscr{L}} \times \mathbb{Z}^+ \to 2^{\mathscr{L}}$ is a function for updating KB that models the pattern of inference of the reasoner. A fact about the pattern of inference of a reasoner is that the reasoner may perform different inferences from the same premises. In the model, this fact is expressed using a function π with a numeric parameter (integer) in addition to the parameter for KB. In this context, $\pi(\mathtt{KB}, 1)$ models an inference from KB, $\pi(\mathtt{KB}, 2)$ models another inference from KB, etc. Then function π determines a reasoning sequence $\mathtt{KB}_0, \mathtt{KB}_1, \ldots, \mathtt{KB}_i, \ldots$, where $\mathtt{KB}_0 = \mathtt{KB}$ is the initial epistemic situation and $\mathtt{KB}_{i+1} = \pi(\mathtt{KB}_i, i+1)$. Supposing that the numeric parameter models some order of intention, the reasoning sequence of a reasoner models how the reasoner would reason from the available information if it had enough cognitive resources (e.g. memory, time for reasoning).[23] The explicit beliefs of the reasoner (beliefs) are the sentences in the $\mathtt{KB}_i$. The stable beliefs of the reasoner (beliefs$_\omega$) are the sentences in $\mathtt{KB}_\omega = \bigcup_i \bigcap_{j \geq i} \mathtt{KB}_j$.

A sentence is in $\mathtt{KB}_\omega$ iff there exists an i such that, for all $j \geq i$, $\phi \in \mathtt{KB}_j$. Then ϕ is warranted iff $\phi \in \mathtt{KB}_\omega$. But ϕ is undefeatable by better (further) reasoning iff ϕ is warranted. Then ϕ is undefeatable by better (further) reasoning iff $\phi \in \mathtt{KB}_\omega$. The idea is that ϕ is undefeatable by better (more rational) reasoning iff ϕ is in an ideal $\mathtt{KB}_\omega$ (the $\mathtt{KB}_\omega$ of an ideal$_\omega$ reasoner for $\models^x$).

The definition of an ideal$_\omega$ reasoner for $\models^x$ is the following:

Definition 5 (Ideal$_\omega$ reasoner for $\models^x$) *A reasoner $\mathscr{R} = \langle \mathtt{KB}, \pi \rangle$ is ideal$_\omega$ for $\models^x$ iff*

(i1$_\omega$) $\mathscr{R}$ believes$_\omega$ all and only the logical consequences of its initial epistemic situation ($\mathtt{KB}_0 \models^x \phi \leftrightarrow \phi \in \mathtt{KB}_\omega$);

(i2$_\omega$) $\mathscr{R}$ has a nontrivial set of beliefs$_\omega$ (for some ϕ, $\mathtt{KB}_\omega \not\models^x \phi$).

In this model, the definition of a finite reasoner is the following:

Definition 6 (Finite reasoner) *A reasoner $\mathscr{R} = \langle \mathtt{KB}, \pi \rangle$ is finite only if:*

(f1) All $\mathtt{KB}_i$ *in* $\mathscr{R}$*'s reasoning sequence are finite (finite memory);*

(f2) π *is an effective method for generating each* $\mathtt{KB}_i$ *(finite inferential capacities).*[24]

In this context, Z may be reinterpreted as follows (in terms of beliefs$_\omega$ and ideal$_\omega$ reasoners):

(Z_ω) There exists a finite ideal$_\omega$ reasoner for a nonmonotonic $\models^x$ that fulfills r1-r5 and that reasoner believes$_\omega$ $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$ after some amount of reasoning.

If Z_ω is a reasonable reinterpretation, then p1 and p2 are both true only if Z_ω is true (i.e. p1 $\wedge$ p2 $\rightarrow$ Z_ω). The first step in evaluating Z_ω is to investigate the features of a nonmonotonic $\models^x$ that fulfills r1-r5. The consequence relation $\models^r$ fulfills r1-r4 and $\models^x$ must be nonmonotonic in order to fulfill r5. A natural choice is to investigate a $\models^x$ that is a nonmonotonic extension of $\models^r$.

From now, I will sketch the pattern of reasoning of a finite ideal$_\omega$ reasoner for a nonmonotonic $\models^x$ that fulfills r1-r5 (call it $\mathscr{R}^x_\omega$). Then I will move from semantic to syntactic considerations (see fn. 5). I will suppose that there exists a corresponding nonmonotonic $\models^x$ with sound and complete axiomatization for which $\mathscr{R}^x_\omega$ is ideal$_\omega$.[25] Requirement r3 was the main reason for an ideal reasoner for a deductive $\models^x$ that fulfills r1-r4 not being finite. Then $\mathscr{R}^x_\omega$ must deal with this requirement differently. $\mathscr{R}^x_\omega$ may use the rule of nonconclusive cp and its defeater:

Definition 7 (Nonconclusive cp (ncp)) *Not believing* $\neg\phi$ *(on a priori basis) is nonconclusive reason for believing* $\Diamond\phi$*.*

Definition 8 (Defeater for ncp) *Believing* $\neg\phi$ *(on a priori basis) defeats ncp.*

In order to deal with r5, $\mathscr{R}^x_\omega$ may have the (nonconclusive) rule of statistical syllogisms and its defeater, where a is an arbitrary individual and F and G are arbitrary properties:[26]

Definition 9 (Statistical syllogism (ss)) *If* $r \geq .5$*, then believing* $pr(F|G) > r$ *and* Ga *is nonconclusive reason for believing* Fa *(Pollock 1995, p.68).*[27]

Definition 10 (Defeater for ss) *Believing* Ha *and believing* $pr(F|G\&H) < pr(F|G)$ *defeats the ss (Pollock 1990, p.9).*

In this context, it may be possible to construct a finite $\mathscr{R}^x_\omega$ that is ideal$_\omega$ for a nonmonotonic $\models^x$ that fulfills r1-r4. $\mathscr{R}^x_\omega$ would be constructed in such a way that, at each stage of the reasoning sequence, it attempts to apply every rule in the axiomatization to all possible combinations of sentences in $\mathtt{KB}_i$. The key point is to construct

$\mathscr{R}^x_\omega$ in such a way that, every time it concludes that ϕ using a nonmonotonic rule, it checks, after each subsequent stage whether a defeater was derived. If a defeater was derived, $\mathscr{R}^x_\omega$ withdraws the conclusion that ϕ (along with all conclusions derived from ϕ). There exist two possibilities: either the defeater is derived (and ϕ is deleted) at some stage or the defeater is never derived and ϕ is never deleted. In any case, $\mathscr{R}^x_\omega$ would believe$_\omega$ ϕ iff $\models^x \phi$ and $\mathscr{R}^x_\omega$ would be ideal$_\omega$ for $\models^x$. $\mathscr{R}^x_\omega$ would be finite because all $\mathtt{KB}_i$ in the reasoning sequence are finite ($\mathtt{KB}_\omega$ is not in the reasoning sequence, f1) and no rule in the axiomatization of $\models^x$ depends on checking whether $\not\models^x \phi$ ($\phi \notin \mathtt{KB}_\omega$), but only whether $\phi \notin \mathtt{KB}_i$ (f2).

If the construction of $\mathscr{R}^x_\omega$ is possible and $\mathscr{R}^x_\omega$ believes$_\omega$ $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$ after some amount of reasoning, then p1 and p2 are true. However, the use of ideal$_\omega$ reasoners for $\models^x$ as the model of ideal negative conceivability has some consequences for the negative zombie argument. The first consequence is that whereas the pattern of inference of an ideal reasoner for $\models^x$ is unreachable for finite reasoners, the pattern of inference an ideal$_\omega$ reasoner for $\models^x$ may be seen as an extrapolation of the pattern of inference of a finite reasoner. In this sense, ideal negative conceivability would be a more amenable principle of modal epistemology for finite reasoners. The second consequence is that, in this model, the relation between (ideal negative) conceivability and possibility cannot be of (deductive) entailment. $\mathscr{R}^x_\omega$ believing $\Diamond\phi$ at some stage of a reasoning sequence is, at most, a nonconclusive reason for $\Diamond\phi$ being true. In this context, cp would be a nonconclusive principle of modal epistemology (i.e. ncp) and the negative zombie argument would be not be conclusive. Finally, $\mathscr{R}^x_\omega$ does not believe$_\omega$ $\Diamond(p \wedge \neg q)$ for every choice of q, what I discuss in next section.

4. Something is conscious!

Requirement r5 states that $\models^x$ must be such that, for all superphysical ϕ, $\models^x p \rightarrow \phi$. The exact nature of p is difficult to grasp, but it is usually accepted that some ϕ are superphysical and, consequently, it must be the case that $\models^x p \rightarrow \phi$. For example, it is usually accepted that the truths of neuroscience n (e.g. 'C-fiber is stimulated') are superphysical and, consequently, it must be the case that, for all n, $\models^x p \rightarrow n$, where the n are the truths of neuroscience.

What about truths of the form 'the probability of q is x' ($pr(q) = x$) and 'n and q have a correlation of x' ($corr(n, q) = x$), where n is a sentence of neuroscience, q is a phenomenal sentence, and x is a number? Are those truths superphysical? I think that this must be the case if these sentences are interpreted as using the notion of natural or nomological probabilities, which are defined in terms of (proportions of) physically possible worlds (see Pollock 1990). The idea is that p contains the funda-

mental physical laws, which determine the physically possible worlds and possible worlds are saturated entities. Then, for each possible world w, either q is true at w or q is false at w (and the same holds for n). Then there exists a fixed number that is the proportion of the physically possible worlds where q is true in relation to all physically possible worlds (and the same holds for n). Then, for any $pr(q) = x$, the value of x is determined by p. And, for any $corr(n,q) = x$, the value of x is determined by p. A change on these values would entail a change in the (proportions of) physically possible worlds, and, consequently, in the fundamental physical laws, and, consequently, in p. Then, in this interpretation, for all q and n, there exists an x such that $pr(q) = x$ and $corr(n,q) = x$ are superphysical and, consequently, it must be the case that $\models^x p \rightarrow pr(q) = x$ and $\models^x p \rightarrow corr(n,q) = x$. Chalmers accepts that the relevant interpretation of probabilities is that using (proportions of) physically possible worlds.[28] In the following, I assume these sentences are superphysical and that $\models^x$ fulfills r5.

The value of Pearson correlation coefficient for two variables X and Y ($corr(X,Y)$) is the result of dividing the covariance of the variables by the product of their standard deviations:

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

The value of $corr(X,Y)$ is such that $-1 \leq corr(X,Y) \leq 1$. A positive $corr(X,Y)$ indicates that Y tends to increase when X increases, a negative $corr(X,Y)$ indicates that Y tends to decrease when X increases, and a $corr(X,Y) = 0$ indicates that X and Y are uncorrelated. In statistics, an indicator variable is a variable X which may only take two values: $X = 0$ and $X = 1$. For indicator variables, the notion of correlation entails the following:[29]

$$corr(X,Y) > 0 \text{ iff } pr(Y = 1|X = 1) > pr(Y = 1).$$

In a two-valued framework, sentences are indicator variables, where $\neg x$ is the value $X = 0$ and x is the value $X = 1$. Then the above theorem may be rewritten as the following:

$$corr(n,q) > 0 \text{ iff } pr(q|n) > pr(q).$$

I intend to show that, contrary to Chalmers, $p \wedge \neg q$ is not ideally negatively conceivable for every choice of q. In the Z_ω model, if $p \wedge \neg q$ is ideally negatively conceivable for every choice of q, then, for every choice of q, $\mathscr{R}^x_\omega$ does not believe $p \rightarrow q$ after any amount of reasoning.[30] But, I intend to show, $\mathscr{R}^x_\omega$ believes $p \rightarrow q$ after some amount of reasoning for every q such that $pr(q) \geq .5$ and for which there exists a truth n such that $corr(n,q) > 0$. This is the case because n is a truth of neuroscience and

$\models^x p \rightarrow n$ (n is superphysical).[31] Given that $corr(n,q) > 0$, $\models^x p \rightarrow corr(n,q) > 0$ ($corr(n,q) > 0$ is superphysical). But $corr(n,q) > 0$ only if $pr(q|n) > pr(q)$, then $\models^x p \rightarrow pr(q|n) > pr(q)$. Given that $pr(q) \geq .5$, $\models^x p \rightarrow pr(q|n) > .5$ ($pr(q) \geq .5$ is superphysical). Then $\models^x p \rightarrow (n \wedge pr(q|n) > .5)$. Then, applying the statistical syllogism on n and $pr(q|n) > .5$, it follows that $\models^x p \rightarrow q$. Then $\mathscr{R}^x_\omega$ believes $p \rightarrow q$ after some amount of reasoning for every q such that $pr(q) \geq .5$ and for which there exists a truth n such that $corr(n,q) > 0$. (def. 2, i1). Therefore, $\mathscr{R}^x_\omega$ does not negatively conceive $p \wedge \neg q$ for every q such that $pr(q) \geq .5$ and for which there exists a truth n such that $corr(n,q) > 0$.

Empirical data suggest that there exist truths n and q such that $corr(n,q) > 0$. There exists an extensive literature on neural correlates of conscious that supports this claim (see Tononi and Koch 2008, for a review). An anonymous referee has pointed out that some philosophers may worry that the acceptance of these correlations may depend on presuppositions about the results of the dualism-physicalism debate. Chalmers, however, is not among these philosophers:

> We have good reason to believe that subjective experiences are systematically correlated with brain processes and behavior. ...We simply need to distinguish correlation from explanation. Even if first-person data cannot be wholly explained in terms of third-person data, the two sorts of data are still strongly correlated. ...The science of consciousness can remain neutral on these philosophical questions. One can simply regard the principles as principles of correlation while staying neutral on their causal and ontological states. (Chalmers 2010, p.40, p.47)

I think that it is dubious how to establish $pr(q)$ for a given q, but it is very likely that $pr(q) > .5$ for many qs of interest. For example, it is very likely that $pr(q) \approx 1$ for $q =$'something is conscious'. The number humans alive is approximately 7×10^9. Let $q_1 =$'human 1 is conscious' and, more generally, $q_i =$'human i is conscious'. Then $pr(q) = 1 - \prod_{i=1}^{7\times10^9} pr(\neg q_i)$. If the probability of each human being conscious is as low as 1×10^{-9}, then $pr(q) \approx 1$ (.999). If the probability of each human being conscious is 1×10^{-10}, then $pr(q) = .503$ (what is still greater than .5). Then the conditions of the argument in this section are very likely fulfilled for some true n and q. For example, most probably, $\mathscr{R}^x_\omega$ believes $p \rightarrow q$ after some amount of a priori reasoning for $q =$'something is conscious'. Therefore, most probably, $p \wedge \neg q$ is not ideally negatively conceivable (in the Z_ω model) and the negative zombie argument does not work for $q =$'something is conscious'. The same holds for any truth q such that $pr(q) \geq .5$ and for which there exists some truth n such that $corr(n,q) > 0$.

This argument is not affected by Chalmers' structure and dynamics counterargument:

> First, physical descriptions of the world characterize the world in terms of structure and dynamics. Second, from truths about structure and dynamics,

> one can deduce only further truths about structure and dynamics. Third, truths about consciousness are not truths about structure and dynamics. (Chalmers 2010, p.120)

Let's grant that (i) physical truths and laws (e.g. the conjuncts in p) are about structure and dynamics, (ii) phenomenal truths (e.g. q) are not about structure and dynamics, and (iii) there does not exist deduction from sentences about structure and dynamics to sentences that are not about structure and dynamics. Premise iii may be true of deductions (which are valid in virtue of form), but Chalmers' own presuppositions entail that the logic of a priori (modal) knowledge (and, hence, $\models^x$) cannot be deductive and iii is not true of nondeductive arguments. But, since all superphysical truths must follow from p in $\models^x$ (r5), simply stating that q should not follow from p in $\models^x$ without an independent reason would be begging the question.

Other objections may appear. First, it may be questioned whether the statistical syllogism is a rational rule of inference. This objection is not very strong. The statistical syllogism is successfully employed in our everyday and scientific reasoning. More acutely, it may be questioned whether the statistical syllogism is a rational rule of inference for *a priori* reasoning. There exists some discussion about whether the statistical syllogism is a rational rule for a priori reasoning (see Russell 2013). I will not discuss this issue here. My point is that Chalmers' negative zombie argument *requires* something like the statistical syllogism to be a rational rule of inference for a priori reasoning (because of the interpretative principles). Finally, it may be questioned whether the conclusion of the argument is an artifact of a too low value of r (i.e. $r = .5$). It may be the case that the specific line of reasoning presented above depends on this specific value of r, but it is possible to construct similar lines of reasoning for values of r very close to 1 because, for some qs of interest (e.g. $q =$'something is conscious'), $pr(q) \approx 1$.

5. Conclusions

The negative zombie argument is not conclusive, but it may still be a strong nonconclusive argument for some choice of q. The cogency of the argument, however, would depend on which q is chosen. I will not discuss the issue about whether there exists a q that satisfies the negative zombie argument. My point is that this is an empirical question. In order to establish whether some specific q satisfies the negative zombie argument, one needs empirical information for supporting either $pr(q) < .5$ or $\forall n(n \rightarrow corr(n,q) \leq 0)$. But any empirical support for either $pr(q) < .5$ or (more dramatically) or $\forall n(n \rightarrow corr(n,q) \leq 0)$ would itself be nonconclusive. Then the negative zombie argument is neither a priori nor conclusive. The negative zombie argument is not a priori because the choice of an adequate q depends on empirical

information. The negative zombie argument is not conclusive because $\mathscr{R}^{x}_{\omega}$ negatively conceiving $p \wedge \neg q$ is at most nonconclusive reason for $\Diamond(p \wedge \neg q)$ and any empirical support for a q being adequate is nonconclusive.

Chalmers argues that ideal positive conceivability entails ideal negative conceivability:

> Ideal primary positive conceivability entails ideal primary negative conceivability: if S (a sentence) can be ruled out a priori, then no coherent imagined situation will verify S. (Chalmers 2010, p.148)

If this is the case, $p \wedge \neg q$ not being ideally negatively conceivable for an arbitrary choice of q has the consequence of $p \wedge \neg q$ not being ideally positively conceivable for an arbitrary choice of q. In addition, if ideal negative conceivability and ideal positive conceivability are the only two relevant kinds of ideal conceivability, this has the consequence of $p \wedge \neg q$ not being ideally conceivable in general for an arbitrary choice of q. Then, if Chalmers is correct, the conclusions presented here have consequences for zombie arguments in general.

References

Antonelli, A. 2005. *Grounded Consequence for Defeasible Logic*. Cambridge University Press.

Binkley, R. 1968. The Surprise Examination in Modal Logic. *Journal of Philosophy* **65**(5): 127–36.

Bohm, D. 1952. A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables. *Phys. Rev.* **85**(2): 166–79.

Boolos, G.; Burgess, J. P. R.; Jeffrey, C. 2007. *Computability and Logic*. Cambridge University Press.

Brewka, G.; Niemel, I.; Truszczyski, M. 2008. Nonmonotonic reasoning. In: F. van Harmelen; Porter, B. (eds) *Handbook of Knowledge Representation*. Volume 3 of Foundations of Artificial Intelligence, pp.239–84. Elsevier.

Carnap, R. 1946. Modalities and quantification. *The Journal of Symbolic Logic* **11**(2): 33–64.

Chalmers, D. 2002. Conceivability and Possibility. In: *Does Conceivability Entail Possibility?*, pp.145–200. Oxford University Press.

———. 2010. *The Character of Consciousness*. Oxford University Press.

———. 2012. *Constructing the World*. Oxford University Press.

Church, A. 1936. A note on the entscheidungsproblem. *Journal of Symbolic Logic* **1**: 40–41.

Cohnitz, D. 2012. The Logic(s) of Modal Knowledge. In: *New Waves in Philosophical Logic*. Palgrave Macmillan.

Duc, H. N. 1995. Logical omniscience vs. logical ignorance on a dilemma of epistemic logic. *Lecture Notes in Computer Science* 990: 237–248.

Evnine, S. 2008. Modal epistemology: Our knowledge of necessity and possibility. *Philosophy Compass* **3/4**: 664–684.

Gendler, T. S.; Hawthorne, J. 2002. *Conceivability and Possibility*. Oxford University Press.

Giunchiglia, E.; Giunchiglia, F. 2001. Ideal and real belief about belief. *J. Logic Computat* **11**: 157–192.

Grim, P. 1988. Logic and limits of knowledge and truth. *Noûs* **22**(3): 341–367.

Halpern, J. Y.; Moses, Y. 1985. Towards a theory of knowledge and ignorance: Preliminary report. In: *Logics and Models of Concurrent Systems*, pp.459–476. Springer-Verlag.

Haugeland, J. 1982. Weak supervenience. *American Philosophical Quarterly* **19**(1): 93–103.

Kirk, R. 2012. *Zombies*. URL: http://plato.stanford.edu/entries/zombies/.

Kraus, S., Lehman, D., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and comulative logics. Artificial Intelligence, 44(1-2): 167–207.

Kripke, S. 1959. A completeness theorem in modal logic. *Journal of Symbolic Logic* **24**(1): 1–14.

Makinson, D. 1994. General patterns in nonmonotonic reasoning. In: D. Gabbay; C. Hogger; J. Robinson (eds.) *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3*, pp.35–110. Oxford and New York: Oxford University Press.

McLaughlin, B.; Bennett, K. 2014. *Supervenience*. URL: https://plato.stanford. edu/entries/supervenience/.

Menzies, P. 1998. Possibility and Conceivability: A Response-Dependent Account of Their Connections. *European Review of Philosophy*: 255–277.

Parent, T. 2016. An objection to the Laplacean Chalmers. *Journal for General Philosophy of Science* **47**(1): 237–240.

Pollock, J. L. 1990. *Nomic Probability and the Foundations of Induction*. Oxford: Oxford University Press.

———. 1995. *Cognitive Carpentry: a blueprint for how to build a person*. The MIT Press.

Ray, C. 1991. *Time, Space, and Philosophy*. Routledge.

Russell, B. 2013. *A priori justification and knowledge*. URL: http://plato.stanford.edu/entries/apriori/.

Salmon, N. 1989. The logic of what might have been. *Philosophical Review* **98**(1): 3–34.

Stanalker, R. 1994. Nonmonotonic consequence relations. *Fundamenta Informaticæ* **21**: 7–21.

———. 2006. On logics of knowledge and belief. *Philosophical Studies* **128**(1): 169–199.

Tononi, G.; Koch, C. 2008. The neural correlates of consciousness. *Annals of the New York Academy of Sciences* **1124**(1): 239–261.

DANILO FRAGA DANTAS
Department of Philosophy
University of California
Davis, California
dfdantas@ucdavis.edu

Notes

[1] Usually, 'nontrivial modal knowledge' ('trivial' in the sense of banal) refers either to knowledge that $\Box\phi$ or to knowledge that $\Diamond\phi$ which is not based on knowledge that ϕ (see Evnine

2008, p.665). For simplicity, 'nontrivial modal knowledge' refers in this paper only to knowledge that $\Diamond\phi$ when ϕ is not even believed. This will simplify the arguments when dealing with ideal reasoners of a specific kind (see fn. 6).

[2] In general, the $\mathscr{L}$-truths are the truths in $\mathscr{L}$. For example, the physical truths are the true sentences of a(n ideal) physical language (a language in which the physics true of the acutal world can be expressed) (Haugeland 1982, p.96). The *fundamental* physical truths and laws is the smallest set of physical truths and laws such that the set of all physical truths and laws supervenes on that set (see fn. 3). If p must contain "the full information about the distribution of the fundamental physical entities through space and time" (Chalmers 2012, p.xiv), then p must be an infinitary conjunction (see section 2).

[3] A set of sentences $\mathscr{L}$ supervenes on a set of sentences $\mathscr{L}'$ iff two possible worlds cannot be discernible with $\mathscr{L}$ without being discernible with $\mathscr{L}'$, where two worlds are discernible with $\mathscr{L}$ when there exists a sentence in $\mathscr{L}$ which is true of one world and is not true of the other (Haugeland 1982, p.96).

[4] An ideal reasoner is defined along these lines in Binkley (1968), Halpern (1985), Grim (1988), Duc (1995), Giunchiglia and Giunchiglia (2001), and Stalnaker (2006). A trivial set of beliefs is a set that entails all the sentences in the language. If the consequence relation is explosive, then nontriviality is equivalent to consistency.

[5] I use a *semantic* consequence relation because I am concerned with the requirements that an ideal reasoner must fulfill in order to be used in the negative zombie argument. The consequence relation $\models^x$ is used as a parameter in those requirements. A reasoner $\mathscr{R}$ being ideal for $\models^x$ does not mean that $\mathscr{R}$ reasons according to some semantic method related to $\models^x$, but only that $\mathscr{R}$ fulfills requirements i1 and i2 interpreted in terms of $\models^x$. In general, talk about requirements will be semantic and descriptions of patterns of reasoning will be syntactic. In section 3, for example, the description of $\mathscr{R}$'s pattern of reasoning is syntactic.

[6] An anonymous referee has pointed out that 'believes $\Diamond(p \wedge \neg q)$ on the basis of not believing $p \rightarrow q$' is too weak and should be replaced with 'believes $\Diamond(p \wedge \neg q)$ on the basis of not having (or having searched and not found) a priori grounds for believing $p \rightarrow q$'. However, since an ideal reasoner believes all and only the tautologies of $\models^x$ and an adequate $\models^x$ for evaluating the negative zombie argument is 'the logic of a priori (modal) knowledge' (see discussion around footnote 7), these conditions coincide. An ideal reasoner for an adequate $\models^x$ believes ϕ iff it has (searched and found) a priori grounds for believing ϕ iff there are a priori grounds for believing ϕ iff ϕ is a priori.

[7] Chalmers argues that the zombie argument denies an epistemic entailment from the physical to the phenomenal truths, where 'epistemic entailment' is explained as follows: "On this notion, ϕ implies ψ when the conditional 'If ϕ then ψ' is a priori — when a subject can know that if ϕ is the case, then ψ is the case with justification independent of experience" (Chalmers 2010, p.109).

[8] The superphysical truths are the truths that supervene on the fundamental physical truths and laws.

[9] The language of $\models^x$ must contain modal operators because $\Diamond(p \wedge \neg q)$ contains modal operators ($\Diamond$). Also, it must contain quantifiers because physical laws are quantified sentences and p has physical laws as conjuncts.

[10] Suppose that $\models^r \phi$. Then, for all $w \in W$, $w \Vdash \phi$. Since all ws are consistent (possible worlds), it is not the case that, for all w, $w \Vdash \neg\phi$ (in fact, for all w, $w \nVdash \neg\phi$). Then $\not\models^r \neg\phi$.

[11] Consider an arbitrary atomic sentence ϕ and an arbitrary world $w \in W$. There exists a w' such that $w' \Vdash \phi$ because ϕ is atomic and W contains all possible assignments for the atomic sentences. wRw' because w is fully connected. Then $w \Vdash \Diamond\phi$. Since w is arbitrary, $\models^r \Diamond\phi$. Also, there exists a w' such that wRw' and $w \nVdash \phi$ because ϕ is atomic and W contains all possible assignments for the atomic sentences. Then $\not\models^r \phi$. Therefore, $\models^r \Diamond\phi$ and $\not\models^r \phi$.
[12] Suppose that $\not\models^r \neg\phi$. Then there exists a w' such that $w' \nVdash \neg\phi$, what means that $w' \Vdash \phi$ (possible world). For all $w \in W$, wRw' because R is fully connected. Then, for all $w \in W$, $w \Vdash \Diamond\phi$ and $\models^r \Diamond\phi$. Therefore, if $\not\models^r \neg\phi$, then $\models^r \Diamond\phi$.
[13] Suppose that $\models^r \phi$. Then, for all $w \in W$, $w \Vdash \phi$. Then, for all $w \in W$, all the $w' \in W$ such that wRw' are such that $w' \Vdash \phi$. Then, for all $w \in W$, $w \Vdash \Box\phi$ and $\models^r \Box\phi$. Therefore, if $\models^r \phi$, then $\models^r \Box\phi$ (necessitation). Consider an arbitrary w. Suppose that $w \Vdash \phi$. wRw (R is fully connected), then $w \Vdash \Diamond\phi$. Then, for all w, if $w \Vdash \phi$, then $w \Vdash \Diamond\phi$. Then $\models^r \phi \rightarrow \Diamond\phi$ (reflexivity).
[14] A set is recursively enumerable iff there exists a(n effective) procedure that outputs all and only the members of the set in some ordering and executes finitely many operations in generating each output. An ideal reasoner for $\models^x$ believes all and only the sentences in $\models^x$, where $\models^x$ is the set of x's theorems. Then $\models^x$ being recursively enumerable means that there exists a procedure (a pattern of reasoning) from which the reasoner may form each of its beliefs executing at most finitely many operations (inferential steps).
[15] I am supposing that $\models^x$ must be consistent (supraclassicality), but the problem presented here also may be avoided using a paraconsistent $\models^x$ (Cohnitz 2012, p.73).
[16] Let $\top$ be a tautology of $\models^x$. For some ψ, $\not\models^x \psi$ (r2). Then, for some $w \in W$, $w \nVdash \psi$ (otherwise, $\models^x \Box\psi$ and $\models^x \psi$, r4). Then $w \nVdash \psi \wedge \top$. Then, for infinitely many $\top$ (there are infinitely many $\top$, supraclassicality), $w \nVdash \psi \wedge \top$. Let $\phi = \psi \wedge \top$. If $w \nVdash \phi$, then $w \Vdash \neg\phi$ (possible worlds). Therefore, for some $w \in W$ and infinitely many ϕ (there are infinitely many $\top$), $w \Vdash \neg\phi$. Therefore, for infinitely many ϕ, $\models^x \Diamond\neg\phi$.
[17] A set is recursively decidable iff both the set and its complement are recursively enumerable. The set $\models^x$ is enumerable iff it is decidable because enumerating the infinitely many $\Diamond\neg\phi$ in this set is equivalent to enumerate the set $\not\models^x$, the complement of $\models^x$.
[18] There are interpretations under which QM is deterministic (Bohm 1952), but quantum indeterminacy does not need to be actual. Here, it is enough that quantum indeterminacy is possible because $\models^x \phi$ iff $\models^x \Box\phi$.
[19] A consequence relation $\models^x$ is deductive when $\Gamma \models^x \phi$ iff there does not exist a model in which all sentences in Γ are true and ϕ is false, where Γ is a set of sentences in the language of $\models^x$. I am exploiting the relation between $p \models^x \phi$ and $\models^x p \rightarrow \phi$.
[20] A well-behaved nonmonotonic logic has the properties of supraclassicality, reflexivity, cut, and cautious monotony (Antonelli 2005, Stanalker 1994, Makinson 1994). A nonmonotonic logic with those properties has a well-behaved semantic (Kraus et. al. 1990), but is not recursively enumerable (see Brewka et. al. 2008).
[21] The phrase 'undefeatable by *better* reasoning' also has an evaluative character ('better' in the sense of more rational). Later in this section, I will contemplate this character by dealing with ideal reasoners.
[22] Defeaters may be rebutting or undercutting. Rebutting defeaters attack the conclusion of a reason. For example, if perceiving at distance what appears to be a sheep in the field is nonconclusive reason for believing there is a sheep in the field, then hearing from the shep-

herd that there are no sheep in the field is a rebutting defeater for that reason (Pollock 1995, p.85). Undercutting defeaters attack the connection between a reason and its conclusion. For example, if perceiving an object as red is nonconclusive reason for believing the object is red, then learning that the object is under red light is an undercutting defeater for that reason. Undercutting defeaters are reasons for believing that a reason does not support a conclusion (Pollock 1995, p.96).

[23] The counterfactual interpretation is not the only interesting interpretation of a reasoning sequence. There exists, for examples, an interesting actualist interpretation in which a reasoning sequence is a plan about how to reason and the i is a temporal index.

[24] $\mathtt{KB}_i$ is finite iff all sentences in $\mathtt{KB}_i$ are finitary and the number of sentences in $\mathtt{KB}_i$ is finite. That π is an effective method for generating each $\mathtt{KB}_i$ entails that each $\mathtt{KB}_i$ is effectively enumerable. Since $\mathtt{KB}_\omega$ is not among the $\mathtt{KB}_i$, $\models^x$ does not need to be recursively enumerable for the existence of ideal$_\omega$ reasoners for $\models^x$.

[25] It is not obvious that there exists such a $\models^x$ with a sound and complete axiomatization. I will grant this because I want to stress more serious problems for the (negative) zombie argument.

[26] $\mathscr{R}^x_\omega$ may use statistical syllogism for concluding, for example, that a electron is in some region from the fact that a high enough proportion of the (squared) amplitude of its wavefunction is concentrated within in that region (Chalmers' interpretative principles).

[27] These are indefinite probabilities, where $pr(F|G)$ means 'the proportion of physically possible Gs that would be Fs', The value for r must be low enough for fulfilling requirement r5. I will suppose that $r = .5$ (the lowest reasonable value), but the argument in next section holds for values of r very close to 1.

[28] "The sort of possibility being considered here is natural or nomological possibility or possibility compatible with the laws of nature. If we required correlation across all logically possible cases, there might be no total NCC [neural correlate of consciousness] at all, as it is logically possible or coherently conceivable to instantiate any physical process at all without consciousness. If we require correlation across naturally possible cases, the problem goes away, as these cases are probably not naturally possible" (Chalmers 2010, p.74).

[29] The correlation between two random variables X and Y is such that $corr(X,Y) = cov(X,Y)/(\sigma_X\sigma_Y)$ and $cov(X,Y) = E[XY]-E[X]E[Y]$, where $E[X]$ is the expected value of X. If X and Y are indicator variables, $E[X] = pr(X=1)$, $E[Y] = pr(Y=1)$, and $E[XY] = pr(X=1\wedge Y=1)$. Then $corr(X,Y) = [pr(X=1\wedge Y=1)-pr(X=1)pr(Y=1)]/(\sigma_X\sigma_Y)$. Since σ_X and σ_Y are always positive, $corr(X,Y) > 0$ iff $pr(X=1\wedge Y=1)-pr(X=1)pr(Y=1) > 0$. But $pr(X=1\wedge Y=1)-pr(X=1)pr(Y=1) = pr(X=1)pr(Y=1|X=1)-pr(X=1)pr(Y=1) = pr(X=1)[pr(Y=1|X=1)-pr(Y=1)]$ and $pr(X=1)$ is always positive. Then $corr(X,Y) > 0$ iff $pr(Y=1|X=1)-pr(Y=1) > 0$. Therefore $corr(X,Y) > 0$ iff $pr(Y=1|X=1) > pr(Y=1)$.

[30] Since believing ϕ after some amount of (a priori) reasoning is a defeater for negatively conceiving $\neg\phi$ (ncp), if $\mathscr{R}^x_\omega$ believes $p\rightarrow q$ after some amount of (a priori) reasoning, then $\mathscr{R}^x_\omega$ does not negatively conceive $p\wedge\neg q$.

[31] I am supposing that the n and q such that $corr(n,q) = x$ are about the same individual. For example, $n = N(a)$ and $q = Q(a)$, where a is some individual.