UvA-DARE (Digital Academic Repository)

Nonmonotonicity and Knowability: As Knowable as Possible

van Rooij, R.

DOI

[10.1007/978-3-319-47843-2_3](https://doi.org/10.1007/978-3-319-47843-2_3)

Publication date

2017

Document Version

Author accepted manuscript

Published in

Rohit Parikh on Logic, Language and Society

[Link to publication](#)

Citation for published version (APA):

van Rooij, R. (2017). Nonmonotonicity and Knowability: As Knowable as Possible. In C. Başkent, L. S. Moss, & R. Ramanujam (Eds.), *Rohit Parikh on Logic, Language and Society* (pp. 53-65). (Outstanding Contributions to Logic; Vol. 11). Springer.
https://doi.org/10.1007/978-3-319-47843-2_3

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Nonmonotonicity and Knowability: As Knowable as Possible

Robert van Rooij
Institute for Logic, Language and Computation

1 Introduction: The knowability paradox

In recent years, Fitch's knowability paradox received a lot of attention. According to some, this paradox proves that verificationism, or more generally, Dummett's anti-realistic theory of meaning is impossible. According to others, there must be something wrong with the proof: how else can it be explained that Dummett's reasonable type of anti-realism reduces to absurdities like naive idealism and even lingualism?

According to anti-realism, if it is impossible to know something, this something cannot be true. In formal terms, $\neg\Diamond\mathbf{K}\phi \rightarrow \neg\phi$. Contrapositively, this means that if something is true, it is knowable, i.e., it can be known:

- $\phi \rightarrow \Diamond\mathbf{K}\phi$. Anti-realism

The problem is that this seemingly innocent assumption gives rise to trouble when combined with some standard assumptions concerning the behavior of the involved modal operators ' $\Diamond$ ' (or its dual ' $\Box$ ') and ' $\mathbf{K}$ '. The standard assumptions are the following:

$$\begin{aligned} \mathbf{N} \quad & \models \phi \quad \Rightarrow \quad \models \Box\phi \\ \mathbf{K} \quad & \models \mathbf{K}(\phi \rightarrow \psi) \rightarrow (\mathbf{K}\phi \rightarrow \mathbf{K}\psi). \\ \mathbf{T} \quad & \models \mathbf{K}\phi \rightarrow \phi. \end{aligned}$$

From **N** and **K** one can straightforwardly prove that **C**, i.e. $\models \mathbf{K}(\phi \wedge \psi) \rightarrow \mathbf{K}\phi \wedge \mathbf{K}\psi$ holds. Let us now assume that we know that there are some unknown truths: $\exists\phi(\mathbf{K}(\phi \wedge \neg\mathbf{K}\phi))$. Let p be one of those. Thus $\mathbf{K}(p \wedge \neg\mathbf{K}p)$. The trouble is that from this assumption together with anti-realism and the standard modal assumptions **N**, **K**, and **T**, one can derive the seemingly absurd thesis that any truth is already known.

To show this, assume $\mathbf{K}(p \wedge \neg\mathbf{K}p)$. With **C** it follows that $\mathbf{K}p \wedge \mathbf{K}\neg\mathbf{K}p$, and thus $\mathbf{K}p$. But from the other conjunct $\mathbf{K}\neg\mathbf{K}p$ it follows with **T** that $\neg\mathbf{K}p$. Thus, from $\mathbf{K}(p \wedge \neg\mathbf{K}p)$ one can derive a contradiction, which means (according to classical logic) that $\mathbf{K}(p \wedge \neg\mathbf{K}p)$ cannot be true, i.e. $\models \neg\mathbf{K}(p \wedge \neg\mathbf{K}p)$. By the necessitation rule **N** it follows that $\Box\neg\mathbf{K}(p \wedge \neg\mathbf{K}p)$, which is equivalent to $\neg\Diamond\mathbf{K}(p \wedge \neg\mathbf{K}p)$.

According to anti-realism, all truths are knowable, formalized as $\phi \rightarrow \Diamond\mathbf{K}\phi$. Now take ϕ to be our assumption that p is an unknown truth: $\phi \equiv p \wedge \neg\mathbf{K}p$. With anti-realism it would follow that $\Diamond\mathbf{K}(p \wedge \neg\mathbf{K}p)$. But from our above reasoning we have concluded that $\neg\Diamond\mathbf{K}(p \wedge \neg\mathbf{K}p)$, which states the exact opposite. We must conclude that according to the anti-realist there can be no proposition p that is true

but unknown $\neg\exists p(p \wedge \neg\mathbf{K}p)$. Stating it otherwise, this leads with **T** to the seemingly absurd proposition that truth and knowledge are indistinguishable, $\forall p(p \leftrightarrow \mathbf{K}p)$: **idealism**.

It is perhaps worth stating that the reasoning behind the paradox leads to **lingualism** as well, i.e. the thesis that all truths are truly expressed, $\forall p(p \leftrightarrow \mathbf{E}p)$, at least on the mild assumption that the factive modality of ‘truly expressed’, **E**, satisfies axiom **K** too. Suppose that we assume *the thesis of expressibility*: that if something is true, it can be expressed, and thus be truly expressed: $\phi \rightarrow \Diamond\mathbf{E}\phi$, for any ϕ . If we now make the natural assumption that the sentential operator **E** is a necessity operator and validates both (i) $\mathbf{E}\phi \rightarrow \phi$ and (ii) $\mathbf{E}(\phi \wedge \psi) \rightarrow \mathbf{E}\phi \wedge \mathbf{E}\psi$, we can in exact analogy with the derivation of idealism immediately conclude that all truths are truly expressed: *lingualism*. Of course, the predicted requirement for any truth that it has to be expressed seems even more absurd than the prediction of (naive) idealism.

There are some well-known obvious reactions to the knowability paradox. First, one might simply bite the bullet, and accept the idealistic conclusion: there is no p that is both true and unknown. Second, one might argue that the paradox shows that anti-realism, just as the thesis of expressibility, is simply wrong. Both conclusions are radical, far-reaching, and not very popular. Indeed, isn’t it more natural to assume that the absurd consequences show that there is something wrong with the proof?

Indeed, much more popular is Tennant’s (1997) response: anti-realism should be limited to so-called *Cartesian* propositions, i.e., propositions that do not lead to contradictions if anti-realism is assumed. This response is completely correct: full-fledged anti-realism cannot escape the conclusion that some truths lead to inconsistency.¹ According to another response (Beall, 2000), one should not worry too much about our knowledge being contradictory, because a closely related paradox (i.e., Montague’s *knower paradox*) already shows that we do have contradictory knowledge, and such contradictory knowledge need not lead to as much disaster as classical logicians assumed. Also Beall seems to be correct: not allowing knowledge to be self-contradictory requires unnatural limitations on what can be expressed.

Still, even the latter two more moderate responses are not fully satisfying: Tennant’s proposal seems more like an *escape* than a solution to the paradox, and although we agree with Beall’s observation that the *knower*-paradox is relevant to the solution of the knowability paradox, Beall seems to *weaken* the anti-realistic thesis, i.e. $\phi \rightarrow \Diamond\mathbf{K}\phi$ for any ϕ , much more than desired. The reason is that if Beall uses the most popular paraconsistent logic that allows for sentences to be both true and false, i.e. the logic *LP*,² it follows that one can only conclude for any true ϕ that the formula $\Diamond\mathbf{K}\phi$ is also true, but not that $\Diamond\mathbf{K}\phi$ is *only true*, i.e., he allows a version of semantic anti-realism where $\Diamond\mathbf{K}\phi$ is false, for any arbitrary truth ϕ .

In this paper we will nevertheless build on the proposals of Tennant and Beall. We won’t give up on anti-realism for the sake of argument, and we will assume that there are some unknown truths. In contrast to Tennant (1997), we will assume (again for the sake of argument) that the thesis of anti-realism, i.e. $\phi \rightarrow \Diamond\mathbf{K}\phi$, holds for

¹Van Benthem (2004) argues that Tennant’s limitation is too weak. The thesis of verificationism should not be limited to propositions that can be known consistently, but rather to propositions that one can learn consistently. Van Benthem observes that there are sentences that can be known consistently, but that cannot be consistently learned: $\phi = (p \wedge \Diamond\neg p) \vee \mathbf{K}\neg p$.

²*LP* is so named by Priest (1979), who has advanced the logic widely in philosophy, though the logic was introduced, for the same reason of accommodating ‘glutty theories’ (true inconsistent but non-trivial theories), by Asenjo (1966) and Asenjo and Tamburino (1975).

any arbitrary ϕ , thus also for non-Cartesian ones. Like Beall (2000) we will make use of a logic that allows contradictions to be true, but our anti-realism will be stronger than the one stated by Beall: for most propositions ϕ we don't allow that if ϕ is true, $\Diamond \mathbf{K}\phi$ can still be false (as well as true). We will make use of Tennant's distinction between Cartesian and non-Cartesian propositions, and propose that for all Cartesian propositions ϕ that are true, $\Diamond \mathbf{K}\phi$ will be *only true*.

What is distinctive about our approach is that we will make use of pragmatic interpretation to account for this notion of 'only truth'. To make use of Tennant's distinction between Cartesian and non-Cartesian propositions, we will make use of a 3-valued semantics, and a non-monotonic consequence relation. I will discuss the prospects of using the non-monotonic logic LP^m that is based on LP to strengthen Beall's (2000) thesis of anti-realism. I will conclude that the logic won't do, because it would lead to the undesired consequence that there can be no unknown truths after all. In the next sections I will motivate the three-valued logic used by Cobreros et al (2012, 2013), and introduce a new non-monotonic logic that accounts for conversational implicatures and for the pragmatic interpretation of vague language. Afterwards I will use these notions to do the required work about knowability.

2 3-valued semantics & pragmatic interpretation

2.1 3-valued semantics

Let $\mathcal{M} = \langle \mathcal{D}, \mathcal{I} \rangle$ be a three-valued model for predicate logic, where D is a domain, as usual, and where $\mathcal{I}$ is a total function from atomic sentences to $\{0, 1, \frac{1}{2}\}$. Now we can define the truth values of sentences as follows:³

- $\mathcal{V}_{\mathcal{M}}(\phi) = \mathcal{I}_{\mathcal{M}}(\phi)$, if ϕ is atomic
- $\mathcal{V}_{\mathcal{M}}(\neg\phi) = 1 - \mathcal{V}_{\mathcal{M}}(\phi)$
- $\mathcal{V}_{\mathcal{M}}(\phi \wedge \psi) = \min\{\mathcal{V}_{\mathcal{M}}(\phi), \mathcal{V}_{\mathcal{M}}(\psi)\}$
- $\mathcal{V}_{\mathcal{M}}(\phi \vee \psi) = \max\{\mathcal{V}_{\mathcal{M}}(\phi), \mathcal{V}_{\mathcal{M}}(\psi)\}$
- $\mathcal{V}_{\mathcal{M}}(\forall x\phi) = \min\{\mathcal{V}_{\mathcal{M}}([\frac{d}{x}]\phi) : d \in D\}$

We say that ϕ is **strictly true** in $\mathcal{M}$ iff $\mathcal{V}_{\mathcal{M}}(\phi) = 1$, and that ϕ is **tolerantly true** iff $\mathcal{V}_{\mathcal{M}}(\phi) \geq \frac{1}{2}$. In terms of this semantics we can define some well-known logics: Kleene's $K3$ and LP . According to both logics, the consequence-relation is truth preserving. The only difference between the two is that while according to $K3$ only value 1 counts as true, according to LP , both 1 and $\frac{1}{2}$ do (while in $K3$ value $\frac{1}{2}$ stands for 'neither true nor false', in LP it denotes 'both true and false'). Thus, $\Gamma \models^{K3} \phi$ iff $\forall \mathcal{M} : \text{if } \forall \gamma \in \Gamma : \mathcal{V}_{\mathcal{M}}(\gamma) = 1, \text{ then } \mathcal{V}_{\mathcal{M}}(\phi) = 1$, and $\Gamma \models^{LP} \phi$ iff $\forall \gamma \in \Gamma : \mathcal{V}_{\mathcal{M}}(\gamma) \geq \frac{1}{2}, \text{ then } \mathcal{V}_{\mathcal{M}}(\phi) \geq \frac{1}{2}$. In some recent joint publications with Pablo Cobreros, Paul Egré, and David Ripley, we showed that a slight variant of $K3$ and LP can account for paradoxes of vagueness (Cobreros et al (2012) and transparent truth (Ripley 2012, Cobreros et al, to appear) in an, arguably, more satisfying way than either $K3$ or LP can. The crucial idea of the analysis of vagueness and transparent truth is that, although we don't give up the idea that entailment is truth-preserving, we allow the 'strength' of truth of the conclusion to be weaker than the strength of

³Notice that the semantics is just like that of Fuzzy Logic, but now limited to three truth values.

truth of the premisses. We say that a sentence ψ is *st-entailed* by a set of premisses Γ , $\Gamma \models^{st} \phi$, iff $\forall \mathcal{M} : \text{if } \forall \gamma \in \Gamma : \mathcal{V}_{\mathcal{M}}(\gamma) = 1$, then $\mathcal{V}_{\mathcal{M}}(\phi) \geq \frac{1}{2}$. This analysis has two immediate consequences: (i) it interprets value $\frac{1}{2}$ as a notion of truth, just like *LP* does, and it thus allows for certain sentences to be both true *and* false, (ii) the notion of consequence is *non-transitive*. One appealing feature of the logic is that in contrast to either *K3* and *LP*, it is a conservative extension of classical logic: it only differs from classical logic if we extend the language with (i) a similarity relation ' $\sim$ ' (Cobreros et al, 2012), so that the tolerance principle $(\forall x, y((Px \wedge x \sim_P y) \rightarrow Py))$ becomes valid, or with a truth-predicate ' $\mathcal{T}$ ' (Cobreros et al, to appear) that behaves fully transparent ($\mathcal{V}_{\mathcal{M}}(\mathcal{T}(\phi)) = \mathcal{I}_{\mathcal{M}}(\phi)$ for any ϕ). And even in these cases the resulting logical differences are minimal, it is only in very special cases (i.e., when it gives rise to paradox) that transitivity fails. To illustrate this for vagueness, if we assume that $Px \wedge x \sim_P y \wedge y \sim_P z$ is strictly true, we can conclude using $\models^{st}$ that Py is at least tolererantly true. And if $Py \wedge y \sim_P z$ is, or were, strictly true, we could conclude with $\models^{st}$ that Pz would be at least tolerantly true. However, the two inferences cannot be joined together to give rise to the Sorites paradox: We cannot conclude from the strict truth of $Px \wedge x \sim_P y \wedge y \sim_P z$ to the tolerant truth of Pz .

2.2 Borderline contradictions

In Cobreros et al (2012) we argued that if Adam is a borderline case of a tall man, the sentence 'Adam is tall' is both true and false. We motivated this by a number of recent experiments (i.e., Alxatib and Pelletier, 2011; Ripley, 2011) that show that naive speakers find a logical contradiction like 'Adam is tall and Adam is not tall' acceptable exactly in case Adam is a borderline tall man. In Cobreros et al (2012) we proposed that the explanation is that we always interpret a sentence pragmatically in the strongest possible way. This pragmatic interpretation accounts, on the one hand, for the intuition that if one *says* that Adam is tall, what is *meant* is that Adam is *only* tall, but, on the other, for the experimentally observed acceptability of contradictions at the border, because contradictions like ' $Ta \wedge \neg Ta$ ' can only be interpreted as true when tolerant truth is at stake. In Cobreros et al (2012) we show that such a pragmatic interpretation also accounts for the observed unacceptability (cf. Serchuk et al., 2011) of classical tautologies like ' $Ta \vee \neg Ta$ ' if Adam is borderline tall. Unfortunately, the interpretation rule gives rise to trouble for more complex sentences. Alxatib, Pagin, and Sauerland (2013) show that we wrongly predict that a sentence like 'Adam is tall and not tall, or John is rich' (of the form $(p \wedge \neg p) \vee q$) means that John is strictly rich, although it intuitively should mean that *either* Adam is borderline tall *or* John is strictly rich. Similarly, our analysis mispredict that a sentence like 'Adam is tall and Adam is not tall, and John is rich' (of the form $(p \wedge \neg p) \wedge q$) can be appropriately asserted if John is not strictly rich.

2.3 Pragmatic interpretation

How should we account for pragmatic interpretation such that we can show what we cannot express, i.e. that a sentence is *only true*, and can solve the above problems with complex sentences involving borderline contradictions? Here is a proposal: the pragmatic interpretation of ϕ makes (exactly) one minimal truth-maker of ϕ as true as possible. What are the minimal truth-makers of ϕ ? and how to think of 'as true as possible'? As for the first question, a minimal truth-maker of ϕ will be thought

of as a set of literals, and the set of minimal truth-makers of ϕ , $T(\phi)$, can easily be defined recursively as follows:⁴

- $T(\phi) = \{\{\phi\}\}$, if ϕ is a literal.
- $T(\phi \vee \psi) = T(\phi) \cup T(\psi)$.
- $T(\phi \wedge \psi) = T(\phi) \otimes T(\psi) = \{A \cup B \mid A \in T(\phi), B \in T(\psi)\}$.
- $T(\forall x \phi) = \bigotimes_{d \in D} T(\phi[x/a])$

Notice that according to these rules, $T(p) = \{\{p\}\}$, $T(\neg p) = \{\{\neg p\}\}$, $T(p \vee q) = \{\{p\}, \{q\}\}$, $T(p \vee \neg p) = \{\{p\}, \{\neg p\}\}$, $T(p \wedge \neg p) = \{\{p, \neg p\}\}$, $T((p \wedge \neg p) \vee q) = \{\{p, \neg p\}, \{q\}\}$, $T((p \wedge \neg p) \wedge q) = T(p \wedge (\neg p \wedge q)) = \{\{p, \neg p, q\}\}$, and $T((p \vee q) \wedge (r \vee s)) = \{\{p, r\}, \{p, s\}, \{q, r\}, \{q, s\}\}$.

To make sense of the notion ‘as true as possible’, we can define an ordering $<_S$ between models, with S a set of sentences (a minimal truth maker), defined as follows: $\mathcal{M} <_S \mathcal{N}$ iff_{def} $\{\phi \in S : \mathcal{V}_{\mathcal{M}}(\phi) = 1\} \subset \{\phi \in S : \mathcal{V}_{\mathcal{N}}(\phi) = 1\}$. In terms of this we define the pragmatic interpretation of ϕ , $PRAG(\phi)$ (where $\llbracket S \rrbracket^t$ abbreviates $\bigcap_{\psi \in S} \llbracket \psi \rrbracket^t$, and where $\llbracket \psi \rrbracket^t = \{\mathcal{M} \mid \mathcal{V}_{\mathcal{M}}(\psi) \geq \frac{1}{2}\}$):

- $PRAG(\phi) = \{\mathcal{M} \in \llbracket S \rrbracket^t \mid S \in T(\phi) \ \& \ \neg \exists \mathcal{N} \in \llbracket S \rrbracket^t : \mathcal{M} <_S \mathcal{N}\}$.

Notice that for literals and conjunctive sentences, this pragmatic interpretation rule simply tries to make its minimal truth-maker as true as possible, i.e., strictly true. Thus, even if one cannot really *express* in language that a sentence like ‘Adam is tall’ is only true, it follows from the above pragmatic interpretation. But in general a sentence might have more than one minimal truth-maker, i.e., when the sentence is disjunctive: so the general pragmatic interpretation rule says that it is enough if one of its minimal truth-makers is as true as possible. As a result, $PRAG((p \wedge \neg p) \wedge q)$ singles out those models where p is tolerantly but not strictly true, *and* where q is strictly true, while $PRAG((p \wedge \neg p) \vee q)$ singles out those models where either p is tolerantly but not strictly true, *or* where q is strictly true.

Of course, the above interpretation rules cannot interpret all sentences translated in a propositional language. For one thing, propositional logic also contains (bi)conditionals, for another, it might be that a negation has scope over a complex sentence. However, both problems can be solved easily, either by also defining the minimal falsemakers of a sentence,⁵ or by stipulating that the pragmatic interpretation rule should be applied to a sentence only after it is put in its so-called ‘Negative Normal Form’.⁶

Let us see what this pragmatic analysis predicts for some examples involving vagueness: (i) ‘ p ’ is interpret as being only true; (ii) ‘ $p \vee \neg p$ ’ is pragmatically interpreted as saying that either p is strictly true, or $\neg p$ is; (iii) ‘ $p \wedge \neg p$ ’ is predicted to be interpreted as saying that p is only tolerantly true; (iv) ‘ $(p \wedge \neg p) \vee q$ ’ is meant as saying that either p is only tolerantly true, or q is strictly true, while (v) ‘ $(p \wedge \neg p) \wedge q$ ’ is predicted to be interpreted as saying that p is only tolerantly true, and that q is strictly true. All these predictions seems to be in accordance with the experimental results and intuition.

⁴The definition of $T(\phi)$ is taken from the analysis of minimal truth-makers proposed by van Fraassen (1969).

⁵Cf. Van Fraassen (1969).

⁶ ϕ' is a sentence in Negative Normal Form of ϕ if (i) it is logically equivalent with ϕ , and (ii) negations only stand in front of atomic formulas.

3 The Knowability paradox in light of the Knower

In Ripley (2012) and Cobreros et al (to appear) it is shown that the logic $\models^{st}$ can account for the liar paradox, in case the language allows for self-reference and has a truth-predicate. But it is easy to show that the same logic can also account for Montague's *knower-paradox*, which is standardly taken to show the limitations of syntactic treatments of modality.⁷ Suppose that we allow in our language both a knowledge-operator and for self-reference. Then we can express a sentence κ (known as *the knower*) that says of itself that it is unknown: $\kappa = \text{this sentence is not known}$. $\kappa := \neg \mathbf{K}(\kappa)$.

Assume that the law of excluded middle, LEM, is valid. Then $\mathbf{K}(\kappa) \vee \neg \mathbf{K}(\kappa)$. If $\mathbf{K}(\kappa)$, then κ (by factivity of $\mathbf{K}$, i.e. $\mathbf{T}$), and by the meaning of κ thus $\neg \mathbf{K}(\kappa)$. Obviously, if $\neg \mathbf{K}(\kappa)$, then $\neg \mathbf{K}(\kappa)$. In either case $\neg \mathbf{K}(\kappa)$, and thus κ . But this means that κ is a validity, $\models \kappa$. By necessitation (rule $\mathbf{N}$) it follows that $\mathbf{K}(\kappa)$. Thus, there is a sentence, namely, κ , for which we can derive both $\mathbf{K}(\kappa)$ and $\neg \mathbf{K}(\kappa)$. By classical logic, it follows that we can derive everything: $\top \models \mathbf{K}(\kappa) \vee \neg \mathbf{K}(\kappa)$; $\mathbf{K}(\kappa) \vee \neg \mathbf{K}(\kappa) \models \mathbf{K}(\kappa) \wedge \neg \mathbf{K}(\kappa)$, and $\mathbf{K}(\kappa) \wedge \neg \mathbf{K}(\kappa) \models \perp$, the logic becomes trivial.

As always, there are various options open to solve this paradox. According to one option, we allow some sentences to be both true and false. If so, there are still various ways to go: either give up on explosion (as in LP), or give up on transitivity of consequence (e.g. by using our $\models_T^{st}$). In both cases, the only way to give a satisfactory semantics for κ is to say that it has value $\frac{1}{2}$. In fact, it has to have this value in all models/worlds. By the meaning of κ , this means that $\mathbf{K}(\kappa)$ has to have value $\frac{1}{2}$ in all models/worlds. What does this mean for the semantics of $\mathbf{K}(\phi)$? The most natural suggestion is to say the following (with $R_{\mathcal{M}}$ as the epistemic accessibility relation):

- $\mathcal{V}_{\mathcal{M},w}(\mathbf{K}(\phi)) = 1$ iff $\forall v \in R_{\mathcal{M}}(w) : \mathcal{V}_{\mathcal{M},v}(\phi) = 1$
- $\mathcal{V}_{\mathcal{M},w}(\mathbf{K}(\phi)) = 0$ iff $\exists v \in R_{\mathcal{M}}(w) : \mathcal{V}_{\mathcal{M},v}(\phi) = 0$
- $\mathcal{V}_{\mathcal{M},w}(\mathbf{K}(\phi)) = \frac{1}{2}$ otherwise.

Notice that this means that $\mathcal{V}_{\mathcal{M},w}(\mathbf{K}(\phi)) = \frac{1}{2}$ iff (i) $\forall v \in R_{\mathcal{M}}(w) : \mathcal{V}_{\mathcal{M},v}(\phi) \geq \frac{1}{2}$ and (ii) $\exists v \in R_{\mathcal{M}}(w) : \mathcal{V}_{\mathcal{M},v}(\phi) \neq 1$. Indeed, on this semantics it comes out that not only κ , but also both $\mathbf{K}(\kappa)$ and $\neg \mathbf{K}(\kappa)$ and thus $\mathbf{K}(\kappa) \vee \neg \mathbf{K}(\kappa)$ and $\mathbf{K}(\kappa) \wedge \neg \mathbf{K}(\kappa)$ have value $\frac{1}{2}$ in any world in any model. It is easy to see that one can solve Montague's knower paradox by giving up explosion. With respect to our consequence-relation $\models^{st}$, however, explosion is valid. But using $\models^{st}$ one can still solve the paradox, because the consequence relation is non-transitive. The tautology $\top$ *st*-entails $\mathbf{K}(\kappa) \vee \neg \mathbf{K}(\kappa)$. But the latter sentence only has value $\frac{1}{2}$. Each disjunct entails κ , and thus it follows that $\mathbf{K}(\kappa) \wedge \neg \mathbf{K}(\kappa)$. The latter, in turn, *st*-entails $\perp$. But nothing disastrous follows, because the two *st*-entailments cannot be put together.

Back to the knowability paradox. We have learned from the knower-paradox that there is at least one instance of ϕ , namely κ , for which we have inconsistent knowledge. Putting it otherwise, we know that the knower is valid, and by its meaning it holds that $\kappa \wedge \neg \mathbf{K}\kappa$ is true. In fact, this sentence *must* have value $\frac{1}{2}$ because both conjuncts have value $\frac{1}{2}$ in all worlds in all models. Let us assume that $\mathcal{V}_{\mathcal{M},w}(\Diamond(\phi)) = 1$ iff $\exists v : \mathcal{V}_{\mathcal{M},v}(\phi) = 1$ and $\mathcal{V}_{\mathcal{M},w}(\Diamond(\phi)) = 0$ iff $\forall v : \mathcal{V}_{\mathcal{M},v}(\phi) = 0$. We will also assume that

⁷‘Thus, if necessity is to be treated syntactically, that is, as a predicate of sentences, as Carnap and Quine have urged, then virtually all of modal logic [...] must be sacrificed’ (Montague, 1963, p. 9).

the strength (strict or tolerant) of $\Box\phi$'s validity in rule **N**, for any necessary-modality '□' (and thus also 'K'), depends on the strength of the validity of ϕ .

$$\mathbf{N} \quad \models^{s/t} \phi \Rightarrow \models^{s/t} \Box\phi$$

But this means that not only $\kappa \wedge \neg\mathbf{K}\kappa$ is tolerantly valid, but that the same holds for $\mathbf{K}(\kappa \wedge \neg\mathbf{K}\kappa)$. In fact, also $\Diamond\mathbf{K}(\kappa \wedge \neg\mathbf{K}\kappa)$ can only have value $\frac{1}{2}$. But from this it follows that the thesis of anti-realism— $\psi \rightarrow \Diamond\mathbf{K}\psi$ —cannot be strictly valid, because $\kappa \rightarrow \Diamond\mathbf{K}(\kappa)$ can only have value $\frac{1}{2}$.⁸

The fact that the thesis of anti-realism cannot be strictly valid does not mean that we have to give up on anti-realism because of the knower paradox: Both *LP* and our logic based on $\models^{st}$ allow us to maintain the validity of $\phi \rightarrow \Diamond\mathbf{K}\phi$, for all ϕ . The reason is that both logics only demand that the conditional is tolerantly true. But notice with Beall (2000) that for exactly the same reason as for *the knower*, the *LP*- and *ST*-validity of $\phi \rightarrow \Diamond\mathbf{K}\phi$ is not a problem for any other sentence p that is (though perhaps only contingently) true and unknown. Our above semantics and interpretation of the thesis of anti-realism allows for $p \wedge \neg\mathbf{K}p$ to even have value 1. The reason is that although we might take the axioms **K** and **T** to hold both strictly and tolerantly,

$$\mathbf{K} \quad \models^{s,t} \mathbf{K}(\phi \rightarrow \psi) \rightarrow (\mathbf{K}\phi \rightarrow \mathbf{K}\psi),$$

$$\mathbf{T} \quad \models^{s,t} \mathbf{K}\phi \rightarrow \phi,$$

from the tolerant truth of $\Diamond\mathbf{K}(p \wedge \neg\mathbf{K}p)$, and thus that of $\mathbf{K}(p \wedge \neg\mathbf{K}p)$ in a world, we cannot even derive the tolerant truth of $\mathbf{K}p \wedge \neg\mathbf{K}p$ in that world, because $\mathcal{V}_{\mathcal{M},w}(\phi \rightarrow \psi) \geq \frac{1}{2}$ if $\mathcal{V}_{\mathcal{M},w}(\phi) = \frac{1}{2}$.

Although both *LP* and *ST* allow us, in contrast to Tennant (1997), to maintain the full thesis of anti-realism, we are still not there where we would like to be. The reason is that we have generalized to the worst case: because there is at least one type of sentence (a sentence of the form $p \wedge \neg\mathbf{K}p$) for which $\phi \rightarrow \Diamond\mathbf{K}\phi$ cannot be strictly true, we cannot demand the thesis to hold strictly for *any* ϕ . What we would like to claim, instead, is that the thesis holds strictly, *except* for those types of sentences for which it gives rise to trouble, i.e., for Tennant's non-Cartesian propositions. It is only natural to try to account for this default behavior in terms of a non-monotonic logic.

4 Anti-realism and non-monotonicity

4.1 Anti-realism and *LP^m*

We have argued above that in case Beall would use *LP* (and makes use of material implication), the thesis of anti-realism would become *too weak*: for any truth ϕ one can conclude at most that $\Diamond\mathbf{K}\phi$ is true, not that it is *only true*. One might try to strengthen this conclusion by making use of a stronger and non-monotonic logic that still allows for contradictions to be true.

There are many paraconsistent logics that are nonmonotonic. One of them is Priest's (1991) *LP^m*, which is strictly stronger than *LP*. Its non-monotonicity is due to minimizing inconsistency: in *LP^m* one looks only at the set of *minimally*

⁸The thesis could still receive value 1, though, if we used Lucasiewicz's conditional, which is like material implication, except that it assigns to $\phi \rightarrow \psi$ value 1 if both ϕ and ψ have value $\frac{1}{2}$.

inconsistent models where the premisses are all tolerantly true. The set of minimally inconsistent worlds/models of ϕ as used in LP^m as $MI(\phi) = \{\mathcal{M} \in \llbracket \phi \rrbracket^t \mid \neg \exists \mathcal{N} \in \llbracket \phi \rrbracket^t : \mathcal{M} < \mathcal{N}\}$, where $\mathcal{M} < \mathcal{N}$ iff_{df} $\{p \in ATOM \mid \mathcal{M} \in \llbracket p \wedge \neg p \rrbracket^t\} \subset \{p \in ATOM \mid \mathcal{N} \in \llbracket p \wedge \neg p \rrbracket^t\}$. If one would use LP to account for the knower and knowability paradox — as suggested by Beall (2000) — it seems that Priest’s (1997) LP^m is a natural candidate to strengthen the force of the anti-realistic thesis as desired. Unfortunately, this candidate won’t do.

An important (and, at least for the cases relevant to this paper, unfortunate) feature of LP^m is that p is an $\models^{LP^m}$ -consequence of $p \vee \phi \dashv p \vee \phi \models^{LP^m} p \dashv$, in case ϕ (but not p) is contradictory. Now consider the thesis of anti-realism: $\phi \rightarrow \Diamond \mathbf{K}\phi$. If ‘ $\rightarrow$ ’ is material implication, this is equivalent to $\neg \phi \vee \Diamond \mathbf{K}\phi$. Thus, using LP^m it follows from anti-realism that we can conclude that ϕ is strictly false, in case $\Diamond \mathbf{K}\phi$ is contradictory. Now take ϕ to be an unknown true proposition, e.g. $p \wedge \neg \mathbf{K}p$. In the introduction we have shown that by standard assumptions of modal logic one can prove that $\neg \mathbf{K}(p \wedge \neg \mathbf{K}p)$ is valid. Using a three-valued logic, one can live with this validity—in combination with the truth of $\Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ —by assuming that the validity of $\neg \mathbf{K}(p \wedge \neg \mathbf{K}p)$ and thus (with **N**) of $\neg \Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ holds only tolerantly: both $\Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ and $\neg \Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ are tolerantly true, which means that $\Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ can at most be both true and false, i.e. is contradictory if true. Let us now try to *use* the validity of the thesis of anti-realism in our reasoning, by making the thesis as strong as possible by looking at the minimally inconsistent models of $(p \wedge \neg \mathbf{K}p) \rightarrow \Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$. Unfortunately, because $\Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ is inconsistent if true, it follows that all minimally inconsistent models where $(p \wedge \neg \mathbf{K}p) \rightarrow \Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ is true are ones where $\neg(p \wedge \neg \mathbf{K}p)$ is strictly true, i.e. ones where the antecedent of the conditional is strictly false! Thus, by using LP^m as our non-monotonic logic in combination with the thesis of anti-realism, it would follow that there can be no unknown truths after all! Even though this idealistic conclusion can only be reached in an empty context, and it might be that $p \rightarrow \neg \mathbf{K}p$ does not hold anymore in case more is assumed (because the logic LP^m is non-monotonic), the fact that idealism is seen as the default case strikes us as unmotivated and undesired.

4.2 A pragmatic consequence relation

The problem with using LP^m as the non-monotonic consequence relation was due to the use of the minimally inconsistent models. Observe that the use of minimally inconsistent models would also predict that a sentence of the form $(p \wedge \neg p) \vee q$ would be interpreted as q . But it was exactly to escape this prediction that we introduced our pragmatic interpretation rule in section 2.3. Now we want to suggest that we should use this pragmatic interpretation to define a new non-monotonic consequence relation in terms of which we can strengthen the thesis of anti-realism.

There are various ways in which we can use our pragmatic interpretation rule to define a non-monotonic consequence relation. Two natural candidates are the following (where $PRAG(\Gamma)$ abbreviates $\bigcap_{\phi \in \Gamma} PRAG(\phi)$ and $PRAG(\Gamma, c)$ abbreviates $\bigcap_{\phi \in \Gamma} PRAG(\phi) \cap c$):

- $\Gamma \models^{prt} \psi$ iff $PRAG(\Gamma) \subseteq \llbracket \psi \rrbracket^t$
- $\Gamma \models_c^{prpr} \phi$ iff_{df} (for all c) $PRAG(\Gamma, c) \subseteq PRAG(\psi, PRAG(\Gamma))$

According to both consequence relations, we should look at the pragmatic interpretation of the premisses. As a result, using neither of the consequence relations

we can conclude q from $(p \wedge \neg p) \vee q$. By looking at the pragmatic interpretation of the premisses, the new relations differ crucially from the consequence relation $\models^{st}$ that we introduced in Cobreros et al (2012). With ' $\models^{st}$ ' as our consequence relation, explosion (from a contradiction, everything can be derived) is predicted to be valid, even if the contradiction was meant to be expressed: $Ta \wedge \neg Ta$ can be interpreted as meaning that Adam is borderline tall. One unfortunate consequence of this picture was that the relation between assertion and inference was lost. Some of what was said (or, better, meant) was ignored in determining what can be inferred from it. The new consequence relations score better on this: If one claims that Adam is borderline tall by saying that he is tall *and* not tall, what one claims is taken seriously.

Another and related unfortunate consequence of using $\models^{st}$ as our consequence relation was that we were forced to make a distinction between the Sorites reasoning with and without the tolerance principle as explicit premise. Without the principle as explicit premiss we predicted in Cobreros et al (2012) that although each step in the argument is valid, the argument as a whole is *invalid*, because the arguments cannot be joined together. We felt, and still feel, that this is intuitively the correct diagnosis of the Sorites paradox. However, in Cobreros et al (2012) we had to claim that with the tolerance principle as explicit premiss, the argument is valid, but that one of the premisses (i.e., the tolerance principle) is not true enough to be used as a premiss in an $\models^{st}$ inference. Making use of one of the new consequence-relations, we can also diagnose the Sorites reasoning with the tolerance principle as explicit premiss as invalid, even though all the steps are valid. The fact that the tolerance principle $\forall x_i, x_j (Px_i \wedge x_i \sim_P x_j \rightarrow Px_j)$ (with $1 \leq i, j \leq n$) cannot be strictly true if both Pa_1 and $\neg Pa_n$ are taken as premisses that are strictly true, does not rule out that it can be used appropriately in an inference where the premisses are interpreted pragmatically.

According to both notions of entailment conjunction-elimination and disjunction-introduction are valid. The fact that we now look at what was meant by a sentence means that, even though $\phi \wedge \neg \phi \models^{prt} \phi$, it does not hold that $\phi \wedge \neg \phi \models^{prt} \psi$. Thus, explosion is not valid. In this sense, both consequence-relations are a type of paraconsistent entailment relations.

Interestingly, our new consequence relations are *non-monotonic*. For instance it will be the case on either consequence-relation that the Disjunctive Syllogism — $p, \neg p \vee q \models q$ — is valid when p and q are atomic sentences, but when we strengthen the first premiss from p to $(p \wedge \neg p)$ the conclusion fails to hold. Moreover, for the first consequence relation $\models^{prt}$, but not for the second ($\models^{prpr}$) it can be that although $\phi \models^{prt} \chi$, it holds that $\phi \wedge \psi \not\models^{prt} \chi$. This can be illustrated by the following example: $Tall(a) \wedge a \sim_T b \models^{prt} Tall(b)$ but $Tall(a) \wedge a \sim_T b \wedge \neg Tall(a) \not\models^{prt} Tall(b)$.

4.3 Pragmatic entailment and anti-realism

We naturally propose that instead of using LP^m to strengthen the thesis of anti-realism, we should make use of one of our pragmatic consequence-relations $\models^{prt}$ or $\models^{prpr}$.⁹ Let us pretend that we can express the anti-realistic thesis in the object language by one sentence ϕ : $\phi \equiv \forall p(p \rightarrow \Diamond Kp)$. Notice that only by using the second consequence relation we can assume that the thesis of anti-realism is valid. However,—and what we feel is more important—on both accounts we can take the

⁹Recall that $\models^{prt}$ is neither transitive (if the language contains a similarity-relation or a transparant truth-predicate) nor does it validate explosion. It is by the latter feature, and not by non-transitivity, that it solves the *the knower*.

thesis of anti-realism as a substantial premise in our reasoning. Taking this thesis as a premiss in reasoning results in the following set of models:

$$\bullet \text{ PRAG}(\phi) = \{\mathcal{M} \in \llbracket S \rrbracket^t \mid S \in T(\phi) \ \& \ \neg \exists \mathcal{N} \in \llbracket S \rrbracket^t : \mathcal{M} <_S \mathcal{N}\}.$$

Notice that for arbitrary p for which $\Diamond \mathbf{K}p$ can be ‘only true’ (i.e., not also false) in some worlds, it is predicted that $\Diamond \mathbf{K}p$ is ‘only true’, in case p is true. This is the same as what would have been predicted by using LP^m , instead. But by using $\models^{prt}$ we now don’t predict anymore that $p \rightarrow \neg \mathbf{K}p$ is the default case for any arbitrary p . By any logic that accepts the thesis of anti-realism and takes the conditional to be material implication, we predict $\neg(p \wedge \mathbf{K}p) \vee \Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ to be true. In contrast to LP^m , though, we don’t predict that thus the first disjunct $p \rightarrow \neg \mathbf{K}p$ must be strictly true, because we allow the second disjunct $\Diamond \mathbf{K}(p \wedge \neg \mathbf{K}p)$ to be tolerantly true. Thus, idealism does not follow as the default case.

References

- [1] Alxatib, S. and F.R. Pelletier (2011), ‘The Psychology of Vagueness: Borderline Cases and Contradictions’. *Mind and Language*. **26**: 287-326.
- [2] Alxatib, S., P. Pagin and U. Sauerland (2013), ‘Acceptable contradictions: pragmatics or semantics? A reply to Cobreros et al.’, *Journal of Philosophical Logic*.
- [3] Asenjo, F. (1966), ‘Calculus of antinomies’, *Notre Dame Journal of Formal Logic*, **7**: 103-105.
- [4] Aseno, F. and J. Tamburino (1975), *Notre Dame Journal of Formal Logic*, **16**: 17-44.
- [5] Beall, Jc. (2006), ‘Fitch’s proof, verificationism, and the knower paradox’, *Australian Journal of Philosophy*, **78**: 2, 241-247.
- [6] Beall, Jc. (2012), ‘Why Priest’s reassurance is not reassuring’, to appear in *Analysis*.
- [7] Benthem, J. van (2004), ‘What one may come to know’, *Analysis*, **64.2**: 95-105.
- [8] Cobreros, P., P. Egge, D. Ripley, and R. van Rooij, (2012), ‘Tolerant, classical, strict’, *Journal of Philosophical Logic*, **41**: 347-385.
- [9] Cobreros, P., P. Egge, D. Ripley, and R. van Rooij, (2013), ‘Reaching transparent truth’, *Mind*, **122**: 841-866
- [10] Fine, K. (1975), ‘Vagueness, truth, and logic’, *Synthese*, **30**, 265-300. 11.
Goodman, N. (1951). *The structure of appearance*. Cambridge: Harvard.
- [11] Fitch, F. (1963), ‘A logical analysis of some value concepts’, *The Journal of Symbolic Logic*, **28**: 135-142.
- [12] Fraassen, B. van (1969), ‘Facts and tautological entailments’, *Journal of Philosophy* **66**: 477-487.

- [13] Grice, H.P. (1967), ‘Logic and conversation’, *William James Lectures*, Harvard University, reprinted in *Studies in the Way of Words*, 1989, Harvard University Press, Cambridge, Massachusetts.
- [14] Montague, R. (1963), ‘Syntactic treatment of modality’, *Acta Philosophica Fennica*, **16**: 153-167.
- [15] Priest, G. (1979), ‘The logic of paradox’, *Journal of Philosophical Logic*, **8**: 219-241.
- [16] Priest, G. (1991), ‘Minimally inconsistent LP’, *Studia Logica*, **50**: 321-331.
- [17] Ripley, D. (2011). ‘Contradictions at the borders’, In R. Nouwen, R. van Rooij, H.-C. Schmitz, & U. Sauerland (Eds.), (2012), *Vagueness in communication*, Berlin: LICS, Springer.
- [18] Salerno, J. (ed.) (2009), *New Essays on the Knowability Paradox*, Oxford: Oxford University Press.
- [19] Serchuk, P., Hargreaves, I., & Zach, R. (2011), ‘Vagueness, logic and use: Four experimental studies on vagueness’, *Mind and Language*, **26**: 540-573.
- [20] Tennant, N. (1997), *The Taming of the True*, Oxford: Oxford University Press.