The Rule of Global Necessitation
SAMUEL A. ALEXANDER
The Ohio State University

Abstract. For half a century, authors have weakened the rule of necessitation in various more or less ad hoc ways in order to make inconsistent systems consistent. More recently, necessitation was weakened in a systematic way, not for the purpose of resolving paradoxes but rather to salvage the deduction theorem for modal logic. We show how this systematic weakening can be applied to the older problem of paradox resolution. Four examples are given: a predicate symbol S4 consistent with arithmetic; a resolution of the surprise examination paradox; a resolution of Fitch’s paradox; and finally, the construction of a knowing machine which knows its own code. We discuss a technique for possibly finding answers to a question of P. Égré and J. van Benthem.

§ 1. Introduction

To make a certain system consistent, Myhill (1960) suggested (pp. 469-470) weakening the rule of necessitation to only range over arithmetical formulas. Since then, others have suggested various other weakenings of necessitation to make other systems consistent. Halbach (2008) restricted necessitation to formulas not involving a particular predicate; Égré (2005) hinted (p. 44) at separating necessitation from soundness. Meanwhile, Fitting (2007) weakened necessitation in a different way, not to repair inconsistencies but instead to recover the deduction theorem for modal logic (see Hakli & Negri (2012) for a survey on this issue). Fitting’s weakening has the advantage that it is very systematic. We apply it to the older objective of fixing inconsistencies.

Suppose we have some axioms which are divided into global and local axioms. The rule of global necessitation for a modal operator \(K\) is as follows:

\[
\frac{\phi}{K(\phi)} \quad \text{provided } \phi \text{ has been proved without appeal to global axioms.}
\]

The rule of global necessitation for a predicate symbol \(K\) is the same, except that \(\phi/K(\phi)\) is replaced by \(\phi/K(\ulcorner \phi \urcorner)\). The idea is that a global axiom is one which is thoroughly trusted by the agent whose knowledge (or belief or...) is represented by \(K\), whereas a local axiom is true, but may not be known by the agent, or may be known with low conviction.

Formally, if a system \(S\) consists of a set of global axioms, a set of local axioms, and the rule of global necessitation for \(K\) (and modus ponens, which we hereafter implicitly include in every system), we say \(S \models \phi\) if there is a sequence \(\phi_1, \ldots, \phi_n\) such that \(\phi_n\) is \(\phi\) and such that for every 1 \(\leq k \leq n\),

1. \(\phi_i\) is a local axiom, or
2. \(\phi_i\) is a global axiom, or
3. \(\phi_i\) is logically valid, or
4. \(\phi_i\) follows from two earlier members of the sequence by modus ponens, or
5. \(\phi_i\) is \(K(\phi_j)\) (or \(K(\ulcorner \phi_j \urcorner)\) if \(K\) is a predicate symbol) for some \(j < i\) such that for every \(1 \leq k \leq j\), \(\phi_k\) is an instance of one of items (2)-(5) of this list.

Received October 2012 1
c⃝ 2012 Association for Symbolic Logic
doi:10.1017/S1755020300000000
For example, the system GLS of provability logic (Boolos (1993) p. 65) is equivalent to a system with global axioms the axioms of GL, local axiom schema T, and the rule of global necessitation.

The idea for this weakened necessitation appeared in Fitting (2007) (p. 94) and more recently it was used by Hakli & Negri (2012). In both cases it was simply called the rule of necessitation; we have called it the rule of global necessitation so as to distinguish it from its stronger ancestor.

The following lemma (compare the suggestion of Smorynski (1984) (p. 454), as described by Hakli & Negri (2012) (p. 854)) is very useful for showing consistency of systems involving the rule of global necessitation. To state it succinctly, we adopt the following notation: if $K$ is a modal operator, $K[\phi]$ will denote $K(\phi)$, and if $K$ is a predicate symbol, $K[\phi]$ will denote $K(⌜\phi⌝)$.

**Lemma 1.1.** Assume a logic where the compactness theorem holds. Suppose $S$ is a system consisting of a set $S_g$ of global axioms, a set $S_\ell$ of local axioms, and the rule of global necessitation for $K$. Let $S'$ be the following system of axioms:

1. $\phi$, if $\phi \in S_g$.
2. $K[\phi]$ if $\phi$ is logically valid.
3. $K[\phi \rightarrow \psi] \rightarrow K[\phi] \rightarrow K[\psi]$.
4. $K[\phi]$ if $\phi$ is an instance of (1)-(3) or (recursively) (4).
5. $\phi$, if $\phi \in S_\ell$.

For any $\phi$, if $S \models \phi$, then $S' \models \phi$.

**Proof.** Let $S_0$ be the system consisting of axioms $S_g$ and the rule of global necessitation. Let $S'_0$ be the set of axioms from lines (1)–(4) of $S'$. We claim: For any $\phi$, if $S_0 \models \phi$ then $S'_0 \models \phi$. This is proved by induction on proof length from $S_0$. The only interesting case is when $\phi$ is $K[\phi_0]$, where $S_0 \models \phi_0$ in fewer steps. By induction, $S'_0 \models \phi_0$. By compactness, there are $s_1, \ldots, s_n \in S_0$ such that

$$s_1 \rightarrow \cdots \rightarrow s_n \rightarrow \phi_0$$

is valid. By (2),

$$S'_0 \models K[s_1 \rightarrow \cdots \rightarrow s_n \rightarrow \phi_0].$$

By repeated applications of (3),

$$S'_0 \models K[s_1] \rightarrow \cdots \rightarrow K[s_n] \rightarrow K[\phi_0].$$

By (4), for each $i$, $K[s_i] \in S'_0$, since each $s_i \in S'_0$ and $S'_0$ is closed under $K$. Thus $S'_0 \models K[\phi_0]$, as desired.

Now we attack the main lemma itself, again by induction on proof length. Suppose $S \models \phi$. There is only one nontrivial case: $\phi$ is $K[\phi_0]$ where $S \models \phi_0$ in fewer steps and $\phi$ is obtained from that shorter proof by the rule of global necessitation. Since $\phi$ is so obtained, this means every step in the proof of $\phi_0$ is an instance of lines (2)-(5) of the definition of proof on the previous page. This implies $S_0 \models \phi_0$, and thus $S_0 \models K[\phi_0]$, and so by the claim, $S'_0 \models K[\phi_0]$, so certainly $S' \models K[\phi_0]$. □

The bulk of the paper will concern modal operator paradoxes. The reason for this is that, due to self-reference, systems involving full necessitation for a predicate
symbol blow up extremely easily
and thus the well-known paradoxes (see Egré (2005)) tend to be low-level, by which I mean they usually derive contradiction
from some set of assumptions which is (equivalent to) a tiny fragment of a predicate-s
ymbol $S4$. Thus, the following theorem more or less resolves them all in one fell
woop (and might, therefore, be a tentative step toward resolving tensions described
in sections 1 & 2 of Halbach & Welch (2009)).

**Theorem 1.2.** (A predicate symbol version of weakened $S4$, consistent with arithmetic) The following system is consistent (in the language of Peano Arithmetic
extended by a predicate symbol $K$):

1. (Global) The axioms of Peano Arithmetic.
2. (Global) $K(⌜φ → ψ⌝) → K(⌜φ⌝ → K(⌜ψ⌝))$.
3. (Global) $K(⌜φ⌝) → K(⌜K(⌜φ⌝)⌝)$.
4. (Local) $K(⌜φ⌝) → φ$.
5. The rule of global necessitation.

(We'll discuss the philosophical plausibility of localizing soundness like this in the Conclusion.)

**Proof.** Let $S'$ be the following system:

1. The axioms of Peano Arithmetic.
2. $K(⌜φ → ψ⌝) → K(⌜φ⌝) → K(⌜ψ⌝)$.
3. $K(⌜φ⌝) → K(⌜K(⌜φ⌝)⌝)$.
4. $K(⌜φ⌝)$ whenever $φ$ is valid.
5. $K(⌜φ⌝)$ whenever $φ$ is an instance of (1)–(4) or (recursively) (5).
6. $K(⌜φ⌝) → φ$.

By Lemma 1.1., we need only show $S'$ is consistent.

In the absence of non-modus ponens rules of inference, consistency is easy to prove: merely construct a model. Let $S'_0$ consist of the axioms in lines (1)-(5) of $S'$.

Let $\mathcal{M}$ be the model which has universe $\mathbb{N}$, which interprets symbols of PA in the intended ways, and which interprets $K$ as follows:

$$\mathcal{M} \models K(⌜φ⌝) \iff S'_0 \models φ$$

(we do not care how $\mathcal{M}$ interprets $K(⌜n⌝)$ if $n$ is not the Gödel number of a formula).

We will show that $\mathcal{M} \models S'$, proving the theorem.

Preliminary Claim: whenever $S'_0 \models φ$, $S'_0 \models K(⌜φ⌝)$. To see this, suppose $S'_0 \models φ$. By compactness, there are $s_1, \ldots, s_n \in S'_0$ such that

$$s_1 → \cdots → s_n → φ$$

is valid. It follows $S'_0 \models K(⌜φ⌝)$ by an argument similar to the proof of Lemma 1.1.

Armed with the claim, we show $\mathcal{M} \models S'$. Suppose $σ \in S'$, we will show $\mathcal{M} \models σ$.

---

1 In Montague (1963) (p. 294) (also quoted in Egré (2005)) we find: “...if necessity is to be treated syntactically ... then virtually all of modal logic, even the weak system $S1$, is to be sacrificed.”

2 One exception is the predicate symbol treatment of the surprise examination paradox, but that paradox has an equally good modal operator treatment anyway. Another exception appears in Horsten & Leitgeb (2001), more on that in our Conclusion below.
Case 1: \( \sigma \) is an axiom of Peano Arithmetic. Then \( \mathcal{N} \models \sigma \) because \( \mathcal{N} \) has universe \( \mathbb{N} \) and interprets symbols of PA in their intended ways.

Case 2: \( \sigma \) is \( K(\forall \phi \rightarrow \psi \gamma) \rightarrow K(\forall \phi \gamma) \rightarrow K(\forall \psi \gamma) \). Assume \( \mathcal{N} \models K(\forall \phi \rightarrow \psi \gamma) \) and \( \mathcal{N} \models K(\forall \phi \gamma) \). By definition this means \( S'_0 \models \phi \rightarrow \psi \) and \( S'_0 \models \phi \). By modus ponens, \( S'_0 \models \psi \), so \( \mathcal{N} \models K(\forall \psi \gamma) \), as desired.

Case 3: \( \sigma \) is \( K(\forall \phi \gamma) \rightarrow K(\forall K(\forall \phi \gamma)) \). Assume \( \mathcal{N} \models K(\forall \phi \gamma) \). This means \( S'_0 \models \phi \). By the preliminary claim, \( S'_0 \models K(\forall \phi \gamma) \), which shows \( \mathcal{N} \models K(\forall K(\forall \phi \gamma)) \).

Case 4: \( \sigma \) is \( K(\forall \phi \gamma) \) where \( \phi \) is valid. Since \( \phi \) is valid, certainly \( S'_0 \models \phi \), so \( \mathcal{N} \models K(\forall \phi \gamma) \).

Case 5: \( \sigma \) is \( K(\forall \phi \gamma) \) where \( \phi \) is an instance of (1)-(5) of \( S' \). Then \( \phi \in S'_0 \), so \( S'_0 \models \phi \), so \( \mathcal{N} \models K(\forall \phi \gamma) \).

Case 6: \( \sigma \) is \( K(\forall \phi \gamma) \rightarrow \phi \). Assume \( \mathcal{N} \models K(\forall \phi \gamma) \), which means \( S'_0 \models \phi \). Note that, by Cases 1–5, \( \mathcal{N} \models S'_0 \). Since \( \mathcal{N} \models S'_0 \) and \( S'_0 \models \phi \), \( \mathcal{N} \models \phi \), completing the proof. \( \square \)

Although we expressed Theorem 1.2 as a result about knowledge, it could just as well be couched as a result about truth predicates (thickening the plot of Friedman & Sheard 1987).

§2. Myhill’s Necessitation and a Moore’s Paradox Myhill (1960) made a certain system consistent by forcing necessitation’s premises to be arithmetical. This differs from our approach: Myhill’s weak necessitation discriminates based on the form of the premise, whereas ours discriminates on the origin of the premise. In this section, we exhibit a paradox thwarting Myhill’s method but repairable with global necessitation.

Let \( \Psi \) be a sentence (in the language of Peano arithmetic) which is true about \( \mathbb{N} \) but independent of PA. For example \( \Psi \) could be \( \text{Con}(PA) \) or \( \Psi \) could be Goodstein’s Theorem (see Kirby & Paris 1982). Thus the Peano arithmetist does not know \( \Psi \), although \( \Psi \) is true.

We might, therefore, attempt to study arithmetists’ knowledge using a system consisting of the axioms of PA, \( \Psi, \neg K(\forall \Psi \gamma) \), and the rule of necessitation. Of course this would be very flawed (and provably inconsistent with a very short proof): the rule of necessitation should not be allowed to interact with \( \Psi \) because the arithmetist does not know \( \Psi \). None of this is surprising, the only reason we bring it up is to point out that Myhill’s weakening of necessitation, restricting it to purely arithmetical premises, has no effect here. On the other hand, the rule of global necessitation can be used to articulate the system properly.

Theorem 2.3. The following system is consistent (\( K \) a predicate symbol):

1. (Global) The axioms of Peano Arithmetic.
2. (Local) \( \Psi \).
3. (Local) \( \neg K(\forall \Psi \gamma) \).
4. The rule of global necessitation.

Though the theorem may rightly be considered silly, the proof introduces a useful trick.

**Proof.** By Lemma 1.1 it suffices to show the following system \( S' \) is consistent:

1. The axioms of Peano Arithmetic.
Global Necessitation

2. $K(\text{"}\phi\text{"})$ whenever $\phi$ is valid.
3. $K(\text{"}\phi \rightarrow \psi\text{"}) \rightarrow K(\text{"}\phi\text{"}) \rightarrow K(\text{"}\psi\text{"})$.
4. $K(\text{"}\phi\text{"})$ whenever $\phi$ is an instance of (1)-(4).
5. $\Psi$.
6. $\neg K(\text{"}\Psi\text{"})$.

Let $S'_0$ be the set of axioms in lines (1)-(4) of $S'$. Let $\mathcal{N}$ be a model with universe $\mathbb{N}$, interpreting symbols of PA in the intended ways, and interpreting knowledge so that

$$\mathcal{N} \models K(\text{"}\phi\text{"})$$

We will show $\mathcal{N} \models S'$, proving the theorem. The tricky part is to show $\mathcal{N} \models \neg K(\text{"}\Psi\text{"})$. For that we need:

Preliminary Claim: $S'_0 \not\models \Psi$.

In the absence of non-modus ponens rules of inference, to show a theory does not prove a formula, it suffices to build a model of the theory where the formula is untrue. Here, it suffices to build a model of $S'_0$ where $\Psi$ is untrue.

Since $\Psi$ is independent of PA, there is a (nonstandard) model $\mathcal{M}$, in the language of PA, such that $\mathcal{M} \models PA$ and $\mathcal{M} \models \neg \Psi$. We will extend $\mathcal{M}$ to a model $\mathcal{M}'$ in the language of PA plus $K$. To do so we must specify how $\mathcal{M}'$ is to interpret $K$.

Let $\mathcal{M}'$ interpret $K$ so that

$$\mathcal{M}' \models K(\text{"}\phi\text{"})$$

(loosely: $\mathcal{M}'$ “knows everything”). I claim $\mathcal{M}' \models S'_0$. To see this, let $\tau \in S'_0$, we will show $\mathcal{M}' \models \tau$.

Case 1: $\tau$ is an axiom of PA. Then $\mathcal{M}' \models \tau$ since $\mathcal{M} \models \tau$ and $\mathcal{M}'$ agrees with $\mathcal{M}$ on the language of PA.

Case 2: $\tau$ is $K(\text{"}\phi\text{"})$ where $\phi$ is valid. Then $\mathcal{M}' \models K(\text{"}\phi\text{"})$ since $\mathcal{M}'$ knows everything.

Case 3: $\tau$ is $K(\text{"}\phi \rightarrow \psi\text{"}) \rightarrow K(\text{"}\phi\text{"}) \rightarrow K(\text{"}\psi\text{"})$. We have $\mathcal{M}' \models K(\text{"}\psi\text{"})$ since $\mathcal{M}'$ knows everything.

Case 4: $\tau$ is $K(\text{"}\phi\text{"})$ where $\phi$ is an instance of (1)-(4). Then $\mathcal{M}' \models K(\text{"}\phi\text{"})$ since $\mathcal{M}'$ knows everything.

Thus $\mathcal{M}' \models S'_0$. And $\mathcal{M} \not\models \Psi$ since $\mathcal{M} \not\models \Psi$ and $\mathcal{M}'$ agrees with $\mathcal{M}$ on the language of PA. This proves the preliminary claim (in fact by arbitrariness of $\Psi$ it proves $S'_0$ is a conservative extension of PA).

For the theorem itself, let $\sigma \in S'$, I claim $\mathcal{N} \models \sigma$.

Case 1: $\sigma$ is an axiom of Peano Arithmetic. Then $\mathcal{N} \models \sigma$ since $\mathcal{N}$ extends $\mathbb{N}$.

Case 2: $\sigma$ is $K(\text{"}\phi\text{"})$ where $\phi$ is valid. Since $\phi$ is valid, $S'_0 \models \phi$, thus $\mathcal{N} \models K(\text{"}\phi\text{"})$.

Case 3: $\sigma$ is $K(\text{"}\phi \rightarrow \psi\text{"}) \rightarrow K(\text{"}\phi\text{"}) \rightarrow K(\text{"}\psi\text{"})$. Assume $\mathcal{N} \models K(\text{"}\phi \rightarrow \psi\text{"})$ and $\mathcal{N} \models K(\text{"}\phi\text{"})$. That means $S'_0 \models \phi \rightarrow \psi$ and $S'_0 \models \phi$. By modus ponens, $S'_0 \models \psi$, so $\mathcal{N} \models K(\text{"}\psi\text{"})$.

Case 4: $\sigma$ is $K(\text{"}\phi\text{"})$ where $\phi$ is an instance of (1)-(4). Then $\phi \in S'_0$, hence $S'_0 \models \phi$ and $\mathcal{N} \models K(\text{"}\phi\text{"})$.

Case 5: $\sigma$ is $\Psi$. We have $\mathcal{N} \models \Psi$ since $\mathcal{N}$ extends $\mathbb{N}$.

Case 6: $\sigma$ is $\neg K(\text{"}\Psi\text{"})$. By the Preliminary Claim, $S'_0 \not\models \Psi$. Thus $\mathcal{N} \not\models K(\text{"}\Psi\text{"})$, as desired, proving the theorem.

Whenever we establish a consistency result by localizing certain schemas, a natural question is what is the minimum set of schemas we can localize and obtain
consistency. It is clear that we cannot strengthen Theorem 2.3 by globalizing $\Psi$. And by L"ob’s Theorem it follows we cannot globalize $\neg K(\Psi)$ either, so in some sense Theorem 2.3 is sharp. Remarkably, though, the same system, formulated using a modal operator instead of a predicate symbol, can be further sharpened.

**Theorem 2.4.** The following system is consistent ($K$ a modal operator):

1. (Global) The axioms of Peano Arithmetic.
2. (Local) $\Psi$.
3. (Global) $\neg K(\Psi)$.
4. The rule of global necessitation.

To prove the theorem, we need a technical lemma which we’ll use again later.

**Lemma 2.5.** Suppose $K$ is a modal operator, $S$ is a set of axioms, and $K(S)$ (the $K$-closure of $S$) is the set of axioms

1. $\phi$, if $\phi \in S$.
2. $K(\phi)$, if $\phi$ is an instance of (1) or (recursively) (2).

Suppose $N$ is a model in which $K$ is interpreted so that $N \models K(\phi)$ iff $N \models \phi$. If $N \models S$ then $N \models K(S)$.

**Proof.** Assume $N \models S$. We prove by induction on formula complexity that for all $\phi \in K(S)$, $N \models \phi$.

Case 1: $\phi$ is of the form $K(\phi_0)$ for some $\phi_0$. If $\phi \in S$ we are done. If $\phi \notin S$, then $\phi$ is an instance of line (2) of $K(S)$, and thus $\phi_0 \in K(S)$. By induction, $N \models \phi_0$, and thus $N \models K(\phi_0)$.

Case 2: $\phi$ is not of the form $K(\phi_0)$. Then $\phi$ must be an instance of line (1) of $K(S)$, i.e., $\phi \in S$, and $N \models \phi$ by assumption.

Like Theorem 2.3, the proof of Theorem 2.4 introduces a trick which we’ll use again later.

**Proof of Theorem 2.4.** By Lemma 1.1 it suffices to show the following system $S'$ is consistent:

1. The axioms of Peano Arithmetic.
2. $\neg K(\Psi)$.
3. $K(\phi)$ whenever $\phi$ is valid.
4. $K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi)$.
5. $K(\phi)$ whenever $\phi$ is an instance of (1)-(5).
6. $\Psi$.

Let $S'_0$ consist of lines (1)-(5). Let $N$ be a model with universe $\mathbb{N}$, interpreting symbols of PA in the intended ways, and interpreting $K$ so that $N \models K(\phi)$ iff $S'_0 \models \phi$.

We’ll show $N \models S'$, proving the theorem. As before, the tricky part is showing $N \models \neg K(\Psi)$. For that we’ll need the same preliminary claim as before, but now with a different proof.

Preliminary Claim: $S'_0 \not\models \Psi$.

To prove $S'_0 \not\models \Psi$ it’s enough to build a model of $S'_0$ where $\Psi$ fails. Since $PA \not\models \Psi$, there is a (nonstandard) model $M$ (in the language of PA) such that $M \models PA$ and
GLOBAL NECESSITATION

\[ M \not\models \Psi. \] Extend \( M \) to a model \( M' \) in the language of PA plus \( K \), by recursively letting

\[ M' \models K(\phi) \iff M' \models \phi \]

for every formula \( \phi \) (this would, of course, be impossible if \( K \) were a predicate symbol). I claim \( M' \models S_0' \). To see this, let \( \tau \in S_0' \), we'll show \( M' \models \tau \).

Case 1: \( \tau \) is an axiom of PA. Then \( M' \models \tau \) since \( M' \models \tau \) and \( M' \) agrees with \( M \) on the language of PA.

Case 2: \( \tau \) is \( \neg K(\Psi) \). By definition, \( M' \models K(\Psi) \iff M' \models \Psi \). And \( M' \not\models \Psi \) since \( M \not\models \Psi \) and they agree on the language of PA. So \( M' \models \neg K(\Psi) \).

Case 3: \( \tau \) is \( K(\phi) \) where \( \phi \) is valid. Since \( \phi \) is valid, \( M' \models \phi \), thus \( M' \models K(\phi) \).

Case 4: \( \tau \) is \( K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi) \). Assume \( M' \models K(\phi \rightarrow \psi) \) and \( M' \models K(\phi) \). This means \( M' \models \phi \rightarrow \psi \) and \( M' \models \phi \). Thus \( M' \models \psi \), so \( M' \models K(\psi) \).

Case 5: \( \tau \) is \( K(\phi) \) where \( \phi \) is an instance of \((1)-(5)\). Then \( M' \models \tau \) by Lemma 2.3.

This establishes \( M' \models S_0' \). And \( M' \not\models \Psi \), since \( M \not\models \Psi \) and the two agree on the language of PA, therefore \( S_0' \not\models \Psi \) and the claim is proved.

Now back to the main theorem, for \( \sigma \in S' \), we must show \( N \models \sigma \). This splits the proof into the same six cases as Theorem 2.3 and the six cases are proved in exactly the same ways as in Theorem 2.3, so we omit them. The crucial difference was the different proof of the Preliminary Claim.

This has an interesting informal application to non-idealized epistemology. As a mathematician, I would like to be able to state the following Moore's paradox:

- Peano Arithmetic (or ZFC) is true,
- The Four-Color Theorem is true (I trust Appel & Haken (1977)),
- I do not know that the Four-Color Theorem is true (I have utmost confidence I will not discover a proof of it in my lifetime),

and operate using the rule of global necessitation (\( \neg K(4CT) \) being global, and 4CT being local). Strictly speaking, this is inconsistent; but we won’t obtain contradiction through purely epistemological means, as the system would be consistent if 4CT were independent. I believe I can work in this system my whole life and not state a contradiction. See Shapiro (2012) p. 7 for a similar example.

The proofs of Theorems 2.3 & 2.4 highlight a difference between operators and predicates: with operators, we can prove \( \not\models \) by constructing toy models where the operator is interpreted as truth, which is impossible with predicates.

§3. Example: The Surprise Examination Paradox The surprise examination paradox hardly needs an introduction (but the survey by Chow (1998) is worthy of pleasure-reading). For a formalization, we turn to McLennan & Chihara (1975), where we find the following system laid out on pp. 76–77 (the language is propositional, with atoms \( p_1,\ldots,p_n \), where \( n \) is the number of days in the week and \( p_i \) is read as “the examination takes place on the \( i \)th day”, further extended by a modal operator \( K \), with \( K(\phi) \) read as “\( \phi \) is known just prior to the exam”):

1. \( K(\phi) \rightarrow \phi \).
2. \( K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi) \).
3. \( T_n \) by which is meant \( p_1 \lor \cdots \lor p_n \) (“an exam will occur”).
4. \( \wedge_{i=1}^n \neg K(p_i) \) (“it will be a surprise”).
The rule of necessitation.

The above system is inconsistent, due to the surprise examination paradox. We would like to make it consistent by weakening necessitation. Following the example in Section 1 in which we made $S4$ consistent for predicate symbols by localizing soundness, we might attempt to resolve the surprise examination paradox in the same way. But it is not difficult to see that the above system remains inconsistent even with $K(\phi) \rightarrow \phi$ localized and necessitation replaced by global necessitation (in fact, the system remains inconsistent even if the $K(\phi) \rightarrow \phi$ schema is completely removed).

It is possible to show the system becomes consistent if we localize both $K(\phi) \rightarrow \phi$ and $\bigwedge_{i=1}^{n} \neg K(p_i)$. This is unsatisfactory, though, because to localize the latter axiom is to suggest the students ignored the “surprise” part of the surprise examination announcement or didn’t take it very seriously. Fortunately, Kritchman & Raz (2011) come to our rescue. When McLelland and Chihara laid down their formulation of the paradox, they did so with the understanding that an event is surprising precisely if it is unknown just prior to its occurrence. Kritchman and Raz point out that, under this definition of surprise, an inconsistent knower is never surprised, since an inconsistent knower knows everything (the same was observed by Halpern & Moses (1986) some time earlier). It seems that if I know everything (including contradictions), that’s as bad as knowing nothing whatsoever. Therefore, we should admit a different definition of surprise:

An event is surprising if either 1) it is unknown just prior to its occurrence, or 2) a contradiction is known.

We know (locally) that the students in the surprise exam paradox are sound, so to us, the two definitions of surprise are equivalent. But the students do not necessarily know as much. Based on this, the axiom $\bigwedge_{i=1}^{n} \neg K(p_i)$ should be weakened to $\bigwedge_{i=1}^{n} \neg K(p_i) \vee K(\bot)$ where $\bot$ is some contradiction (such as $p_1 \land \neg p_1$). Having done so, we can keep it global, obtaining a more satisfactory resolution to the paradox.

**Theorem 3.6.** Assume $n > 1$. The following system is consistent:

1. (Local) $K(\phi) \rightarrow \phi$.
2. (Global) $K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi)$.
3. (Global) $T_n$, by which is meant $p_1 \lor \cdots \lor p_n$.
4. (Global) $\bigwedge_{i=1}^{n} \neg K(p_i) \lor K(\bot)$.
5. (Global) $\bigwedge_{i=1}^{n} ((\neg T_i) \rightarrow K(\neg T_i))$, where $T_i$ abbreviates $p_1 \lor \cdots \lor p_i$.
6. The rule of global necessitation.

**Proof.** By Lemma 1.1. it suffices to prove consistency of the following system $S'$:

1. $K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi)$.
2. $T_n$, by which is meant $p_1 \lor \cdots \lor p_n$.
3. $\bigwedge_{i=1}^{n} \neg K(p_i) \lor K(\bot)$.
4. $\bigwedge_{i=1}^{n} ((\neg T_i) \rightarrow K(\neg T_i))$, where $T_i$ abbreviates $p_1 \lor \cdots \lor p_i$.
5. $K(\phi)$ whenever $\phi$ is valid.
6. $K(\phi)$ whenever $\phi$ is an instance of (1)-(6).
Global Necessitation

7. \( K(\phi) \rightarrow \phi. \)

Let \( S'_0 \) consist of the first six lines of \( S' \). Let \( \mathcal{N} \) be a model where \( p_1 \) is true, \( p_i \) is false for \( i > 1 \), and \( K \) is interpreted so that

\[ \mathcal{N} \models K(\phi) \text{ iff } S'_0 \models \phi. \]

We will show \( \mathcal{N} \models S' \), proving the theorem. To do that, we need a preliminary claim.

Preliminary Claim: For any \( 1 \leq i \leq n \), \( S'_0 \not\models \phi \).

Fix \( 1 \leq i \leq n \). Since \( S'_0 \) involves no non-modus ponens rules, it suffices to build a model of \( S'_0 \) where \( p_i \) fails. Since \( n > 1 \), there is some \( 1 \leq j \leq n \) such that \( j \neq i \). Let \( \mathcal{M} \) be the model where \( p_j \) is true, \( p_k \) is false for all \( k \neq j \), and knowledge is interpreted so that

\[ \mathcal{M} \models K(\phi) \text{ for every } \phi, \]

i.e., \( \mathcal{M} \) knows everything. I claim \( \mathcal{M} \models S'_0 \). Since \( \mathcal{M} \not\models p_1 \), this will prove the preliminary claim. To see \( \mathcal{M} \models S'_0 \), let \( \tau \in S'_0 \), we'll show \( \mathcal{M} \models \tau \).

Case 1: \( \tau \) is \( K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi) \). Well, \( \mathcal{M} \models K(\psi) \) since \( \mathcal{M} \) knows everything.

Case 2: \( \tau \) is \( T_n \). Then \( \mathcal{M} \models \tau \) since \( \mathcal{M} \models p_i \).

Case 3: \( \tau \) is \( (\bigwedge_{i=1}^n \neg K(p_i)) \lor K(\bot) \). Then \( \mathcal{M} \models \tau \) since \( \mathcal{M} \models K(\bot) \).

Case 4: \( \tau \) is \( \bigwedge_{i=1}^n ((\neg T_i) \rightarrow K(\neg T_i)) \). It suffices to show \( \mathcal{M} \models K(\neg T_k) \) for an arbitrary \( 1 \leq k \leq n \). This is true because \( \mathcal{M} \) knows everything.

Case 5-6: \( \tau \) is \( K(\phi) \) where \( \phi \) is valid or \( \phi \) is an instance of (1)-(6). Then \( \mathcal{M} \models K(\phi) \) since \( \mathcal{M} \) knows everything.

This shows \( \mathcal{M} \models S'_0 \), and so since \( \mathcal{M} \not\models p_i \), \( S'_0 \not\models p_i \). The Preliminary Claim is proved.

Back to the main theorem, to show \( \mathcal{N} \models S' \), let \( \sigma \in S' \), we'll show \( \mathcal{N} \models \sigma \).

Case 1: \( \sigma \) is \( K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi) \). Suppose \( \mathcal{N} \models K(\phi \rightarrow \psi) \) and \( \mathcal{N} \models K(\phi) \). This means \( S'_0 \models \phi \rightarrow \psi \) and \( S'_0 \models \phi \). Thus \( S'_0 \models \psi \) and \( \mathcal{N} \models K(\psi) \).

Case 2: \( \sigma \) is \( T_n \). Then \( \mathcal{N} \models \sigma \) since \( \mathcal{N} \models p_1 \).

Case 3: \( \sigma \) is \( (\bigwedge_{i=1}^n \neg K(p_i)) \lor K(\bot) \). To show \( \mathcal{N} \models \sigma \) it suffices to let \( 1 \leq i \leq n \) be arbitrary and show \( \mathcal{N} \models \neg K(p_i) \). By the Preliminary Claim, \( S'_0 \not\models p_i \), thus \( \mathcal{N} \not\models K(p_i) \), as desired.

Case 4: \( \sigma \) is \( \bigwedge_{i=1}^n ((\neg T_i) \rightarrow K(\neg T_i)) \). Then \( \mathcal{N} \models \sigma \) vacuously, because \( \mathcal{N} \not\models \neg T_i \) for any \( i \), because \( \mathcal{N} \not\models p_1 \).

Case 5: \( \sigma \) is \( K(\phi) \) where \( \phi \) is valid. Since \( \phi \) is valid, \( S'_0 \models \phi \), so \( \mathcal{N} \models K(\phi) \).

Case 6: \( \sigma \) is \( K(\phi) \) where \( \phi \) is an instance of (1)-(6). Then \( \phi \in S'_0 \), so \( S'_0 \models \phi \), so \( \mathcal{N} \models K(\phi) \).

Case 7: \( \sigma \) is \( K(\phi) \rightarrow \phi \). Suppose \( \mathcal{N} \models K(\phi) \). This means \( S'_0 \models \phi \). By Cases 1-6, \( \mathcal{N} \models S'_0 \). Since \( \mathcal{N} \models S'_0 \) and \( S'_0 \models \phi \), \( \mathcal{N} \models \phi \) and our work is finished. \[ \square \]

---

3 This technique fails if we want to make \( p_1 \) false in \( \mathcal{N} \), i.e. if we want the surprise examination to fall on a day other than Monday. The proof can be altered, at the price of slightly more complexity, to allow a surprise exam on any day except for day \( n \). For a surprise exam on day \( 1 < k < n \), one would add \( \neg T_{k-1} \) as a global axiom to the system and adjust the proof accordingly, and in the proof of the Preliminary Claim, choose \( j > k \).
§4. Example: Fitch’s Paradox  Fitch’s Paradox, first published in Fitch (1963), is the fact that under certain assumptions, if every true fact is knowable then every true fact is known. See Salerno (2010) for an introduction, though we part ways when it comes to the exact system to use. Instead we turn to Costa-Leite (2006) who obtained the following system by very systematic means (the language is propositional, extended by modal operators $K$ and $\Box$, and (if we understand correctly) $\Diamond$ abbreviates $\neg \Box \neg$):

1. $\Box (\phi \rightarrow \psi) \rightarrow (\Box (\phi) \rightarrow \Box (\psi))$.
2. $(K (\phi \rightarrow \psi) \land K (\phi)) \rightarrow K (\psi)$.
3. $K (\phi) \rightarrow \phi$.
4. The knowability thesis: $\phi \rightarrow \Diamond (K (\phi))$.
5. $\Box$-necessitation: $\phi / \Box (\phi)$.
6. $K$-necessitation: $\phi / K (\phi)$.

From these, the omniscience principle schema, $\phi \rightarrow K (\phi)$, can be deduced by what is known as the Church-Fitch argument. This is considered a paradox since it seems plausible that all truths are knowable ($\phi \rightarrow \Diamond (K (\phi))$) but it seems absurd that we are omniscient. We would like to weaken necessitation to remove this unwanted consequence. It turns out that if we localize $K (\phi) \rightarrow \phi$, it resolves the paradox so strongly, we can even strengthen the knowability thesis.

**Theorem 4.7.** Let $\mathcal{L}$ be a propositional language (with at least one atom $q$) extended by modal operators $K$ and $\Box$. Let $S$ be the following system in $\mathcal{L}$:

1. (Global) $\Box (\phi \rightarrow \psi) \rightarrow (\Box (\phi) \rightarrow \Box (\psi))$.
2. (Global) $(K (\phi \rightarrow \psi) \land K (\phi)) \rightarrow K (\psi)$.
3. (Local) $K (\phi) \rightarrow \phi$.
4. (Global) The strong knowability thesis: $\Diamond (K (\phi))$.
5. The rule of global necessitation for $\Box$.
6. The rule of global necessitation for $K$.

Then $S \not\models q \rightarrow K (q)$.

**Proof.** By a straightforward variation of Lemma 1.1., it suffices to show $S' \not\models q \rightarrow K (q)$, where $S'$ is:

1. $\Box (\phi \rightarrow \psi) \rightarrow (\Box (\phi) \rightarrow \Box (\psi))$.
2. $(K (\phi \rightarrow \psi) \land K (\phi)) \rightarrow K (\psi)$.
3. The strong knowability thesis: $\Diamond (K (\phi))$.
4. $K (\phi)$ whenever $\phi$ is valid.
5. $\Box (\phi)$ whenever $\phi$ is valid.
6. $K (\phi)$ whenever $\phi$ is an instance of (1)-(7).
7. $\Box (\phi)$ whenever $\phi$ is an instance of (1)-(7).
8. $K (\phi) \rightarrow \phi$.

---

4 I should confess here that in Alexander (preprint) I committed a grave sin: I misread Salerno’s system and attributed a different system to him for which he was not responsible.
Global Necessitation

Let $S'_0$ consist of lines (1)-(7) of $S'$ and let $\mathcal{N}$ be a model in which $q$ is true and $K$ and $\Box$ are interpreted so

$$\mathcal{N} \models K(\phi) \text{ iff } S'_0 \models \phi, \text{ and } \mathcal{N} \models \Box(\phi) \text{ iff } S'_0 \models \phi$$

(we admit this is a very unusual semantics for $\Box$, but since the theorem we are proving is entirely syntactical, that should not matter). We will show $\mathcal{N} \models S'$ and $\mathcal{N} \not\models q \rightarrow K(q)$. Since $S'$ involves no non-modus ponens rules of inference, this will show $S' \not\models q \rightarrow K(q)$ and prove the theorem.

Preliminary Claim: There is a model $\mathcal{M}$ such that $\mathcal{M} \models S'_0$, $\mathcal{M} \not\models q$, and $\mathcal{M} \not\models \neg K(\phi)$ for every formula $\phi$.

Let $\mathcal{M}$ be a model in which $q$ is false, $K$ is interpreted so that

$$\mathcal{M} \models K(\phi) \text{ for every } \phi,$$

and $\Box$ is interpreted recursively so that

$$\mathcal{M} \models \Box(\phi) \text{ iff } \mathcal{M} \models \phi.$$

Right away $\mathcal{M} \not\models q$ and $\mathcal{M} \not\models \neg K(\phi)$, it remains to show $\mathcal{M} \models S'_0$. Let $\tau \in S'_0$, we will show $\mathcal{M} \models \tau$.

Case 1: $\tau$ is $\Box(\phi \rightarrow \psi) \rightarrow (\Box(\phi) \rightarrow \Box(\psi))$. The way $\mathcal{M}$ interprets $\Box$, to show $\mathcal{M} \models \tau$ we must show $\mathcal{M} \models (\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \psi)$, but this is a tautology.

Case 2: $\tau$ is $K(\phi \rightarrow \psi) \land K(\phi) \rightarrow K(\psi)$. Then $\mathcal{M} \models \tau$ because $\mathcal{M} \models K(\psi)$.

Case 3: $\tau$ is $\phi(K(\phi))$. Unwrapping the abbreviation, $\tau$ is $\neg\Box(\neg K(\phi))$. By the way $\mathcal{M}$ interprets $\Box$, we must show $\mathcal{M} \models \neg(\neg K(\phi))$, i.e., $\mathcal{M} \models K(\phi)$, which is true by the way $\mathcal{M}$ interprets $K$.

Case 4/5: $\tau$ is $K(\phi)$ or $\Box(\phi)$, where $\phi$ is valid. Trivial.

Case 6: $\tau$ is $K(\phi)$ where $\phi$ is an instance of (1)-(7). Trivial.

Case 7: $\tau$ is $\Box(\phi)$ where $\phi$ is an instance of (1)-(7). Then $\mathcal{M} \models \tau$ by Lemma 2.5.

This shows $\mathcal{M} \models S'_0$ and proves the Preliminary Claim.

Back to the main theorem, we have two things to show: that $\mathcal{N} \not\models q \rightarrow K(q)$, and that $\mathcal{N} \models S'$. Since $\mathcal{N} \models q$, to show $\mathcal{N} \not\models q \rightarrow K(q)$ we must show $\mathcal{N} \not\models K(q)$, i.e., that $S'_0 \not\models q$. But this is true by the Preliminary Claim: there is a model where $S'_0$ holds and $q$ fails.

To show $\mathcal{N} \models S'$, let $\sigma \in S'$, we will show $\mathcal{N} \models \sigma$.

Case 1/2: $\sigma$ is $\Box(\phi \rightarrow \psi) \rightarrow (\Box(\phi) \rightarrow \Box(\psi))$ or $(K(\phi \rightarrow \psi) \land K(\phi)) \rightarrow K(\psi)$. By now, this case should be straightforward.

Case 3: $\sigma$ is $\phi(K(\phi))$. That is, $\sigma$ is $\neg\Box(\neg K(\phi))$. By the Preliminary Claim, there is a structure where $S'_0$ holds and $\neg K(\phi)$ fails. Since there are no non-modus ponens rules of inference in $S'_0$, this shows $S'_0 \not\models \neg K(\phi)$. Thus $\mathcal{N} \not\models \Box(\neg K(\phi))$, as desired.

Case 4/5: $\sigma$ is $K(\phi)$ or $\Box(\phi)$, where $\phi$ is valid. Trivial.

Case 6/7: $\sigma$ is $K(\phi)$ or $\Box(\phi)$, where $\phi$ is an instance of (1)-(7). Then $\phi \in S'_0$, thus $S'_0 \models K(\phi)$ or $\Box(\phi)$.

Case 8: $\sigma$ is $K(\phi) \rightarrow \phi$. Suppose $\mathcal{N} \models K(\phi)$, which means $S'_0 \models \phi$. By Cases 1-7, $\mathcal{N} \models S'_0$. Thus $\mathcal{N} \models \phi$. This proves the theorem.

The above theorem is strictly stronger than the second theorem of Alexander (preprint), in which the knowability thesis was also localized along with $K(\phi) \rightarrow \phi$.

To prove Theorem 4.7, we constructed a model in which $K$ and $\Box$ were interpreted identically. Thus we could strengthen Theorem 4.7 by adding $K(\phi) \leftrightarrow \Box(\phi)$ as a local axiom (call it $K = \Box$). Could we add this axiom globally? The answer
turns out to be “no”; one can see this by using the variation on the Church-Fitch argument which I published in Alexander (2012). It is, however, possible to globalize $K = \Box$ and avoid paradox if, in exchange, the knowability thesis is localized:

**Theorem 4.8.** The following system does not prove $q \rightarrow K(q)$:

1. (Global) $(\Box(\phi \rightarrow \psi) \rightarrow (\Box(\phi) \rightarrow \Box(\psi)))$.
2. (Global) $(K(\phi \rightarrow \psi) \land K(\phi)) \rightarrow K(\psi)$.
3. (Global) $K(\phi) \leftrightarrow \Box(\phi)$.
4. (Local) $K(\phi) \rightarrow \phi$.
5. (Local) The strong knowability thesis: $\Diamond(K(\phi))$.
6. The rule of global necessitation for $\Box$.
7. The rule of global necessitation for $K$.

**Proof Sketch.** Similar to the proof of Theorem 4.7, with $S'$ and $S'_0$ modified in the obvious ways. The crucial difference is the proof of the Preliminary Claim. When constructing the model $\mathcal{M}$ of the Preliminary Claim, make it interpret $K$ and $\Box$ so that

$$\mathcal{M} \models K(\phi) \text{ for all } \phi, \text{ and } \mathcal{M} \models \Box(\phi) \text{ for all } \phi.$$  

Note that $\mathcal{M}$ will then fail the knowability thesis but that is fine, since the knowability thesis is no longer in $S'_0$—but satisfy $K(\phi) \leftrightarrow \Box(\phi)$.

In summary we can say the following about the structure of Fitch’s paradox ($G$ stands for global, $L$ stands for local):

- $G$-Soundness + $L$-Knowability $\models$ Paradox (Fitch (1963)).
- $L$-Soundness + $G$-Knowability $\nless$ Paradox (Theorem 4.7).
- $G$-Knowability + $G$-($K = \Box$) $\models$ Paradox (Alexander (2012)).
- $G$-($K = \Box$) + $L$-Knowability + $L$-Soundess $\nless$ Paradox (Theorem 4.8).

In Halbach (2008) we were asked, “What if the preferred remedy for the paradox of the Knower and other paradoxes of self-reference also helps to resolve Fitch’s paradox?” In light of Theorems 1.2 and 4.7, it seems we’ve achieved precisely that, at least if preferred is read as “preferred by the present author”.

### §5. Example: Machines which know their own codes

It is considered a well-known fact that if a mechanical knowing agent is capable of logic and arithmetic and self-reflection, then that machine cannot know the index of a Turing machine representing it. See Lucas (1961), Benacerraf (1967), Reinhardt (1985), Penrose (1989), Carlson (2000), and Putnam (2006). However, the proofs involve unrestricted necessitation in various guises. We will weaken necessitation and explicitly construct a machine which knows its own code.

Following Carlson (2000), we work in the language of Peano Arithmetic extended by a modal operator $K$ (i.e., the language of Epistemic Arithmetic of Shapiro (1988)). For de re semantics we use Carlson’s base logic. To paraphrase and summarize: a structure for a modal language consists of a first-order structure $\mathcal{N}$ for

---

5 Putnam does not speak so much about arbitrary machines which know their own code, but rather argues that the totality of scientific knowledge is not a machine which knows its own code.
its first-order part, together with a function which takes a purely modal formula $K(\phi)$ and a variable assignment $s$ and outputs True or False—in which case we write $\mathcal{N} \models K(\phi)[s]$ or $\mathcal{N} \not\models K(\phi)[s]$ respectively—satisfying some technical conditions which we will here brush under the rug.

If $\phi$ is a formula with free variables $x_1, \ldots, x_n$ (and no others), a universal closure of $\phi$ is $\forall x_1 \cdots \forall x_n \phi$. To prove that a given model $\mathcal{N}$ satisfies a universal closure of $\phi$, it suffices to let $s$ be an arbitrary assignment and show $\mathcal{N} \models \phi[s]$.

It follows from Carlson (2000) (specifically from Proposition 3.2, Definitions 3.1 and 3.4, and the discussion on p. 54) that the following system is inconsistent (this can be glossed as “a machine cannot know its own code”):

1. $S4$, by which is meant
   (a) Universal closures of $K(\phi \to \psi) \to K(\phi) \to K(\psi)$.
   (b) Universal closures of $K(\phi) \to \phi$.
   (c) Universal closures of $K(\phi) \to K(K(\phi))$.
   (d) The rule of necessitation.

2. The axioms of Epistemic Arithmetic (i.e., Peano Arithmetic, with the induction schema extended to our modal language).

3. (Knowledge of Code) $\exists eK(\forall x(K(\phi) \iff x \in W_e))$, when $x$ is the lone free variable of $\phi$.
   - Here $W_e$ is the $e$th r.e. set and $x \in W_e$ abbreviates a formula stating that $x$ lies therein.

We will localize (1b) and show the resulting system is $\omega$-consistent by explicitly constructing a machine which knows its own code (via Kleene’s Recursion Theorem). Our machine construction method is somewhat similar to the method of Carlson (2000) and Carlson (2012). Whereas my goal is a truthful machine which knows its own code at the cost of its own truth, Carlson’s goal was a truthful machine which knew its own truth and knew it had some code, articulated by Reinhardt’s Strong Mechanistic Thesis: $K(\exists e \forall x(K(\phi) \iff x \in W_e))$. To do this Carlson used a very careful analysis of ordinal arithmetic (Carlson (1999)), later organized into patterns of resemblance in Carlson (2001). The reason Carlson’s goal took so much work was precisely because his machines knew their own truthfulness globally; by localizing that, we’ll have much less difficulty reaching our own goal.

For every $n \in \mathbb{N}$, let $S_n$ be the following system:

1. Weakened $S4$, by which is meant
   (a) (Global) Universal closures of $K(\phi \to \psi) \to K(\phi) \to K(\psi)$.
   (b) (Local) Universal closures of $K(\phi) \to \phi$.
   (c) (Global) Universal closures of $K(\phi) \to K(K(\phi))$.
   (d) The rule of global necessitation.

2. (Global) The axioms of Epistemic Arithmetic (i.e., Peano Arithmetic, with induction schema extended to the modal language).

3. (Global) (Having Code $n$) $\forall x(K(\phi) \iff \langle x, gn(\phi) \rangle \in W_\mathcal{N})$, provided $x$ is the lone free variable in $\phi$.
   - Here $\langle x, gn(\phi) \rangle \in W_\mathcal{N}$ abbreviates a formula saying $\langle x, gn(\phi) \rangle \in W_n$ where $gn(\phi)$ is the Gödel number of $\phi$ and $(\cdot, \cdot) : \mathbb{N}^2 \to \mathbb{N}$ is a canonical computable pairing function.
My goal is to show that there is some \( n \in \mathbb{N} \) such that \( S_n \) is \( \omega \)-consistent. Assuming I can do that much, we can combine global necessitation with (3) to see

\[
S_n \models K(\forall x(K(\phi) \leftrightarrow \langle x, ^\gamma \phi \rangle) \in W_n),
\]

which implies \( S_n \models \exists e K(\forall x(K(\phi) \leftrightarrow x \in W_e)) \).

With Lemma 1.1. in mind, for each \( n \in \mathbb{N} \), let \( S'_n \) be the system

1. Universal closures of \( K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi) \).
2. Universal closures of \( K(\phi) \rightarrow K(\phi) \).
3. The axioms of Epistemic Arithmetic.
4. \( \forall x(K(\phi) \leftrightarrow \langle x, ^\gamma \phi \rangle) \in W_n \) (\( x \) the lone free variable of \( \phi \)).
5. Universal closures of \( K(\phi) \) whenever \( \phi \) is valid (true in every structure according to the base logic).
6. \( K(\phi) \) whenever \( \phi \) is an instance of (1)-(6).
7. Universal closures of \( K(\phi) \rightarrow \phi \).

**Lemma 5.9.** For each \( n \in \mathbb{N} \), and each sentence \( \phi \), if \( S_n \models \phi \), then \( S'_n \models \phi \).

**Proof.** Similar to the proof of Lemma 1.1. (Carlson points out that the base logic satisfies the compactness theorem). \( \square \)

Thus, we only need show that \( S'_n \) is \( \omega \)-consistent for some \( n \in \mathbb{N} \).

If \( \phi \) is a formula and \( s \) is an assignment, let \( \phi^s \) denote the sentence

\[
\phi(x_1|s(x_1))(x_2|s(x_2))\ldots
\]

obtained by replacing every free variable in \( \phi \) by the numeral for its \( s \)-value.

For every \( n \in \mathbb{N} \), let \( S'_{n0} \) be the following set of axioms:

1. Lines (1)-(6) of \( S'_n \).
2. (Assigned Validity) \( \phi^s \), for valid \( \phi \) and \( s \) an assignment.
3. \( K(\phi) \) for each instance of (1)-(3) of \( S'_{n0} \).

**Lemma 5.10.** There is a total computable function \( f : \mathbb{N} \rightarrow \mathbb{N} \) such that for every \( n \in \mathbb{N} \),

\[
W_{f(n)} = \{ \langle m, gn(\phi) \rangle \in \mathbb{N} : \phi \text{ is a formula with } FV(\phi) \subseteq \{x\} \text{ and } S'_{n0} \models \phi(x|\overline{m}) \}
\]

where \( \phi(x|\overline{m}) \) is the result of substituting the numeral \( \overline{m} \) for \( x \) in \( \phi \).

**Proof.** By the Church-Turing Thesis (Carlson points out that the base logic satisfies the completeness theorem; combined with compactness, this implies that the set of consequences of each r.e. theory is r.e., and the code of the set of consequences of a theory depends uniformly computably on the code of the theory). \( \square \)

**Corollary 5.11.** There is an \( n \in \mathbb{N} \) such that

\[
W_n = \{ \langle m, gn(\phi) \rangle \in \mathbb{N} : \phi \text{ is a formula with } FV(\phi) \subseteq \{x\} \text{ and } S'_{n0} \models \phi(x|\overline{m}) \}
\]

**Proof.** By Kleene’s Recursion Theorem and Lemma 5.10. \( \square \)

**Theorem 5.12.** Let \( n \) be as in Corollary 5.11. Then \( S_n \) is \( \omega \)-consistent. In other words, there is a knowing machine which knows that it has code \( n \).
Proof. As discussed above, it suffices to show $S'_{n_0}$ is $\omega$-consistent. Let $\mathcal{N}$ be a structure, in the language of Epistemic Arithmetic, with universe $\mathbb{N}$, which interprets symbols of PA in the intended ways, and which interprets knowledge according to

$$\mathcal{N} \models K(\phi)[s] \text{ iff } S'_{n_0} \models \phi^s.$$ 

We will show that $\mathcal{N} \models S'_{n_0}$. First we need a couple preliminary claims.

Claim 1: For each formula $\phi$ and assignment $s$, $\mathcal{N} \models \phi[s]$ iff $\mathcal{N} \models \phi^s$.

By induction on formula complexity. The only interesting case is when $\phi$ is $K(\phi_0)$. If $\mathcal{N} \models K(\phi_0)[s]$, by definition this means $S'_{n_0} \models \phi_0^s$. Since $\phi_0^s$ is a sentence, for every assignment $t$ we have $(\phi_0^s)^t = \phi_0^t$ and thus $S'_{n_0} \models (\phi_0^s)^t$. This shows that for every such $t$, $\mathcal{N} \models K(\phi_0)[t]$. By arbitrariness of $t$, this shows $\mathcal{N} \models K(\phi_0)$. Clearly $K(\phi_0^s) \equiv K(\phi_0)^s$. The reverse direction—that if $\mathcal{N} \models K(\phi_0)^s$ then $\mathcal{N} \models K(\phi_0)[s]$—is similar.

Claim 2: For each sentence $\phi$, if $S'_{n_0} \models \phi$, then $S'_{n_0} \models K(\phi)$.

By compactness, if $S'_{n_0} \models \phi$, there are $\sigma_1, \ldots, \sigma_\ell \in S'_{n_0}$ such that

$$\sigma_1 \rightarrow \cdots \rightarrow \sigma_\ell \rightarrow \phi$$

is valid. By (5) of $S'_{n_0}$,

$$S'_{n_0} \models K(\sigma_1 \rightarrow \cdots \rightarrow \sigma_\ell \rightarrow \phi).$$

By repeated application of (1) of $S'_{n_0}$,

$$S'_{n_0} \models K(\sigma_1) \rightarrow \cdots \rightarrow K(\sigma_\ell) \rightarrow K(\phi).$$

Since each $\sigma_i \in S'_{n_0}$, each $K(\sigma_i) \in S'_{n_0}$ since $S'_{n_0}$ is closed under $K$. By modus ponens, $S'_{n_0} \models K(\phi)$, as desired.

Armed with these claims, we are ready to show $\mathcal{N} \models S'_{n_0}$. Let $\sigma \in S'_{n_0}$, we will show $\mathcal{N} \models \sigma$.

Case 1: $\sigma$ is a universal closure of $K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi)$. Let $s$ be an assignment and assume $\mathcal{N} \models K(\phi \rightarrow \psi)[s]$ and $\mathcal{N} \models K(\phi)[s]$. This means $S'_{n_0} \models (\phi \rightarrow \psi)^s$ and $\sigma \models \phi^s$. Clearly $(\phi \rightarrow \psi)^s \equiv \phi^s \rightarrow \psi^s$, so by modus ponens, $S'_{n_0} \models \psi^s$, so $\mathcal{N} \models K(\psi)[s]$ as desired.

Case 2: $\sigma$ is a universal closure of $K(\phi) \rightarrow K(K(\phi))$. To get $\mathcal{N} \models \sigma$, let $s$ be an arbitrary assignment and assume $\mathcal{N} \models K(\phi)[s]$. This means $S'_{n_0} \models \phi^s$. By Claim 2, $S'_{n_0} \models K(\phi^s)$, and clearly $K(\phi^s) \equiv K(\phi)^s$, so $S'_{n_0} \models K(\phi)^s$. This shows $\mathcal{N} \models K(K(\phi))[s]$, as desired.

Case 3: $\sigma$ is an axiom of Epistemic Arithmetic. If $\sigma$ is a basic axiom (such as $\forall x(S(x) \neq 0)$ or $\forall x \forall y(x \cdot S(y) = x \cdot y + x)$), then $\mathcal{N} \models \sigma$ simply because $\mathcal{N}$ has universe $\mathbb{N}$ and interprets symbols of PA in the intended ways. But suppose $\sigma$ is a universal closure of an instance

$$\phi(x[0]) \rightarrow (\forall x(\phi \rightarrow \phi(x[S(x)]))) \rightarrow \forall x \phi$$

of the induction schema ($\phi$ is allowed to involve $K$). Let $s$ be an assignment and assume $\mathcal{N} \models \phi(x[0])[s]$ and $\mathcal{N} \models \forall x(\phi \rightarrow \phi(x[S(x)]))[s]$. We must show $\mathcal{N} \models \forall x \phi^s[s]$. Since $\mathcal{N} \models \phi(x[0])[s]$, it follows by Claim 1 that $\mathcal{N} \models \phi^s(x[0])$. Clearly $\phi^s(x[0]) \equiv \phi^{s(x[0])}$, where $s(x[0])$ is the assignment which agrees with $s$ except that it maps $x$ to 0. So $\mathcal{N} \models \phi^{s(x[0])}$.

For each $m \in \mathbb{N}$, since $\mathcal{N} \models \forall x(\phi \rightarrow \phi(x[S(x)]))[s]$, in particular $\mathcal{N} \models \phi \rightarrow \phi(x[S(x)][s(x[m])]. And thus if $\mathcal{N} \models \phi[s(x[m])$, then $\mathcal{N} \models \phi(x[S(x)][s(x[m])].$
By Claim 1, that last sentence can be rephrased: if $\mathcal{N} \models \phi^s(x|m)$, then $\mathcal{N} \models \phi(x|S(x))^s(x|m)$, but clearly $\phi(x|S(x))^s(x|m) \equiv \phi^s(x|m+1)$, so in summary so far:

- $\mathcal{N} \models \phi^s(x|0)$.
- For each $m \in \mathbb{N}$, if $\mathcal{N} \models \phi^s(x|m)$, then $\mathcal{N} \models \phi^s(x|m+1)$.

Therefore, by mathematical induction, $\mathcal{N} \models \phi^s(x|m)$ for every $m \in \mathbb{N}$. This shows (via Claim 1) that for every $m \in \mathbb{N}$, $\mathcal{N} \models \phi[s(x|m)]$. This means precisely that $\mathcal{N} \models \forall x \phi[s]$, as desired.

Case 4: $\sigma$ is $\forall x (K(\phi) \leftrightarrow \langle x, \bar{x} \phi \bar{\gamma} \rangle \in W_{\pi})$, where $x$ is the lone free variable in $\phi$. To show $\mathcal{N} \models \sigma$ it suffices to let $m \in \mathbb{N}$ be arbitrary and prove $\mathcal{N} \models K(\phi) \leftrightarrow \langle x, \bar{x} \phi \bar{\gamma} \rangle \in W_{\pi}[s]$ where $s$ is an assignment with $s(x) = m$. The following are equivalent:

\[
\begin{align*}
\mathcal{N} & \models K(\phi)[s] \\
S'_{m0} & \models \phi^s \quad \text{(Definition of $\mathcal{N}$)} \\
S'_{m0} & \models \phi(x|m) \quad \text{(Since $\phi$ has lone free variable $x$)} \\
\langle m, gn(\phi) \rangle & \in W_n \quad \text{(By choice of $n$ (Corollary 5.11))} \\
\mathcal{N} & \models \langle m, \bar{x} \phi \bar{\gamma} \rangle \in W_{\pi} \quad \text{(Since $\mathcal{N}$ has standard first-order part)} \\
\mathcal{N} & \models \langle (x, \bar{x} \phi \bar{\gamma}) \in W_{\pi}[s] \quad \text{(Since $s(x) = m$)} \\
\mathcal{N} & \models \langle x, \bar{x} \phi \bar{\gamma} \rangle \in W_{\pi}[s] \quad \text{(By Claim 1)}
\end{align*}
\]

as desired.

Case 5: $\sigma$ is a universal closure of $K(\phi)$ where $\phi$ is valid. Let $s$ be any assignment. Since $\phi$ is valid, $\phi^s$ is an instance of the Assigned Validity schema from the definition of $S'_{m0}$. Thus $S'_{m0} \models \phi^s$, hence $\mathcal{N} \models K(\phi)[s]$. Since $\phi$ is valid, $\phi^s$ is a sentence (note: this is where we finally reap our reward for putting all those universal closures everywhere). So for any assignment $s$, $\phi \equiv \phi^s$, and so, since (1)-(6) of $S'_{m0}$ are included in $S'_{m0}$, $S'_{m0} \models \phi^s$, whence $\mathcal{N} \models \phi[s]$.

Case 7: $\sigma$ is a universal closure of $K(\phi) \rightarrow \phi$. Let $s$ be an assignment and assume $\mathcal{N} \models K(\phi)[s]$ (so $S'_{m0} \models \phi^s$); we must show $\mathcal{N} \models \phi[s]$.

First I claim that $\mathcal{N} \models S'_{m0}$. To see this, let $t \in S'_{m0}$, we'll show $\mathcal{N} \models t$.

Subcase 1: $t$ is an instance of (1)-(6) of $S'_{m0}$. Then $\mathcal{N} \models t$ by Cases 1–6.

Subcase 2: $t$ is $\psi^t$ where $\psi$ is valid and $t$ is an assignment. Since $\psi$ is valid, $\mathcal{N} \models \psi[t]$. By Claim 1, $\models \psi^t$.

Subcase 3: $t$ is $K(\psi)$ where $\psi$ is an instance of (1)-(3) of $S'_{m0}$. Similar to Case 6 above.

This shows $\mathcal{N} \models S'_{m0}$. Since $\mathcal{N} \models S'_{m0}$ and $S'_{m0} \models \phi^s$, $\mathcal{N} \models \phi^s$. By Claim 1, $\mathcal{N} \models \phi[s]$, as desired. The theorem is proved. □

What this section has established is that there is a certain dichotomy in machine knowledge:

A truthful machine can know its own code, or it can know its own truthfulness, but not both.

In future work we will explore formulations of Theorem 5.12 in which $K$ is a predicate symbol rather than a modal operator. By localizing soundness, we rule out the self-referential paradoxes which would normally make predicate symbols
Global Necessitation

unimaginable in this role. See Halbach & Welch (2009) for a discussion of the relative merits of operators vs. predicates.

§6. On a question of Égré and van Benthem Ad the end of Égré (2005), the following question appears (attributed to J. van Benthem):

“...from what we saw, schematic Löb’s theorem appears as the positive counterpart of a negative result ... Could there be other Löb-style strengthenings of the inconsistency results we have presented, that provide us with positive information about consistent subsystems of the conflicting schemata?”

Imagine we did not know Löb’s theorem. One way we might find it is as follows. In Theorem 1.2. we proved consistency of a weak predicate form of $S_4$, by constructing a specific model. To probe for things that can be consistently added to the system, rather than consider the class of all models of the system (which is very hard to get our hands on), we now have a particular model. Any schema which holds in the particular model must necessarily be consistent with the system, locally; whether or not it is consistent globally is another matter, which requires special attention. One property we might notice about the particular model is that it satisfies Löb’s theorem. By asking, “can this newfound consistent local axiom be made global,” we might discover Löb’s schema.

More generally, any time we resolve a paradox by the methods of this paper, building a particular model, we can look for new consistent local axioms by simply examining that particular model; we can then attempt to prove (or disprove) they are consistent globally. For example, as seen at the end of Section 4, this kind of reasoning lead us to a new variation of the Church-Fitch argument.

If we find a new globally consistent axiom in this way, it might be a candidate to answer Égré’s and van Benthem’s question. If that axiom somehow explains why a stronger system is inconsistent, even better.

Here are two concrete candidates to answer the question. First, the schema $(\bigwedge_{i=1}^n \neg K(p_i)) \lor K(\bot)$ from Section 3. This author initially resolved the surprise examination paradox by localizing both $K(\phi) \rightarrow \phi$ and $\bigwedge_{i=1}^n \neg K(p_i)$, but this was unsatisfactory; but the schema $(\bigwedge_{i=1}^n \neg (p_i)) \lor K(\bot)$, true in the particular model, could be globalized. The resulting system was a “consistent subsystem of the conflicting schemata”, and the new schema “provides us with positive information” about it.

Likewise the author initially found the schema $\diamondsuit(K(\phi))$ of Section 4 by looking at a specific model witnessing a consistent subsystem of Fitch’s hypotheses, and $\diamondsuit(K(\phi))$ could be argued to provide positive information about that as well.

Along these same lines, by considering the particular model constructed in Section 5, we might notice it satisfies ($x$ the lone free variable in $\phi$)

$$(\forall x(\phi \rightarrow K(\phi))) \rightarrow \exists e \forall x(\phi \leftrightarrow x \in W_e),$$

For another possible answer, see Theorem 2 of Friedman (1975).
and it’s not hard to show this schema can consistently be added to Theorem 5.12 as a global axiom. This schema (and knowledge thereof) is called an \textit{epistemic Church’s Thesis} in Reinhardt (1985b).

§7. Conclusion We have shown that certain paradoxes vanish, and certain impossibilities become possible, by localizing certain axioms. This should not come as too much of a surprise. One might justify necessitation by saying that if we can prove $\phi$, then the knower himself can follow that proof, and therefore knows $\phi$. But for the knower to follow, the proof must only use axioms the knower is aware of. This became very clear in Section 2, where paradox occurred despite Myhill’s weakening of necessitation: we know by second-order logic that Goodstein’s Theorem is true in $\mathbb{N}$, but when we don our first-order Peano arithmetic hats, that knowledge lacks conviction. Merely adding the troublesome axiom would be harmless (as it is true), but its presence invalidates the justification for the rule of necessitation.

In most of our examples, we localized soundness, $K(\phi) \rightarrow \phi$. The fact that this resolved all those paradoxes gives evidence, in our opinion, that we should take greater care before assuming $K(K(\phi) \rightarrow \phi)$. For further evidence, consider the paradox of the knower of Kaplan & Montague (1960) (as described in Anderson (1983) or on p. 21 of ´Egr´e (2005)) where $K(\neg K(\neg \phi))$ is one of just three schemas (plus arithmetic) which lead to paradox, the other two schemas being practically unassailable. And if that is not already a devastating blow against requiring $K(\phi) \rightarrow \phi$, Thomason (1980) proves that the following system proves $K(\neg \phi)$ for every $\phi$ (see p. 23 of ´Egr´e (2005)):

1. $K(\neg \phi) \rightarrow K(\neg K(\neg \phi))$,
2. $K(\neg K(\neg \phi) \rightarrow \phi)$,
3. $K(\neg \phi)$ if $\phi$ is valid,
4. $K(\neg \phi \rightarrow \psi) \rightarrow (K(\neg \phi) \rightarrow K(\neg \psi))$.

This shows that (at least for predicate symbols) mere belief in soundness is already deadly—even without soundness itself!

See Alexander (preprint) for several arguments for the philosophical plausibility of knowledge which fails $K(K(\phi) \rightarrow \phi)$. I became even further convinced of this when I read, in Shapiro (1998), about remarks Gödel made at the Gibbs lecture (Gödel (1951)). The main objection (I think) to admitting the possibility of $\neg K(\phi) \rightarrow \phi$ is that truthfulness is built into the very definition of knowledge. But in his lecture, Gödel distinguished between objective knowledge and subjective knowledge, so that, for example, mathematicians might subjectively come to know that mathematics is sound, through some argument strictly outside mathematics itself— an empirical argument, say. Thus the mathematician might understand, subjectively, that $K(\phi) \rightarrow \phi$ holds for her knowledge, while rejecting $K(K(\phi) \rightarrow \phi)$ (say, upon considering the second incompleteness theorem), not because of doubting the \textit{truth} of $K(\phi) \rightarrow \phi$, but rather because of doubting the \textit{mathematical provability} of it.

Section 5 is particularly relevant: epistemological traditions should assume a back seat when in conflict with practical applications. A truthful knowing machine which

\footnote{And see van Fraassen (2011) for another argument of the same, when $K$ represents belief rather than knowledge.}
knows its own code could be useful in the real world, even without knowing its own truthfulness. We could, perhaps, sidestep the issue by calling such a machine a “believing machine” rather than a “knowing machine”, but this would be disingenuous, because we know the machine is truthful; to call it a believing machine would be to suggest it might be capable of asserting falsities.

Further Work. One obvious direction to take from here is to expand the list of paradoxes treated by global necessitation. One particularly interesting paradox is found on pp. 260–261 of [Horsten & Leitgeb (2001)]. This paradox can be formally treated with the methods of the current paper, but as there are no agents involved in it, philosophical justification will require careful treatment in a future paper.

Having obtained, through global necessitation, consistent variations of certain paradoxical systems, one thing to consider is: which of these consistent systems can be fused while preserving consistency? (Compare the work of Friedman & Sheard (1987) on axioms about self-referential truth.) For example, by merging ideas from the proofs of Theorems 3.36 and 4.77 it is possible to show that if $n > 1$ then the following system is consistent and does not prove $p_1 \rightarrow K(p_1)$, in a language containing operators $K$, $\Box$, and atoms $p_1, \ldots, p_n$, thus “simultaneously resolving” Fitch’s and the Surprise Examination Paradox:

1. (Global) $K(\phi \rightarrow \psi) \rightarrow K(\phi) \rightarrow K(\psi)$.
2. (Global) $T_n$, by which is meant $p_1 \lor \cdots \lor p_n$.
3. (Global) $\bigwedge_{i=1}^n (\neg T_i) \rightarrow K(\neg T_i)$, where $T_i$ abbreviates $p_1 \lor \cdots \lor p_i$.
4. (Global) $(\bigwedge_{i=1}^n \neg K(p_i)) \lor K(\bot)$.
5. (Global) $\Box(\phi \rightarrow \psi) \rightarrow (\Box(\phi) \rightarrow \Box(\psi))$.
6. (Local) The strong knowability thesis: $\Diamond(K(\phi))$.
7. (Local) $K(\phi) \rightarrow \phi$.
8. The rules of global necessitation for $K$ and $\Box$.

Note that in the above system we’ve localized $\Diamond(K(\phi))$; if we kept it global, the same consistency proof would not work (the Preliminary Claims would be hard to reconcile). The consistency or inconsistency of the resulting system is presently unknown to me (even if $\Diamond(K(\phi))$ were weakened to $\phi \rightarrow \Diamond(K(\phi))$).

A similar question can be glossed provocatively:

Is it possible for a truthful knowing machine to know its own code, and simultaneously attend a class in which a surprise examination next week is announced?
Bibliography

Alexander, S. (preprint). An axiomatic version of Fitch’s paradox. To appear in *Synthese*.

Alexander, S. (2012). A purely epistemological version of Fitch’s Paradox. *The Reasoner, 6*, 59–60.

Anderson, C.A. (1983). The paradox of the knower. *The Journal of Philosophy, 80*, 338–355.

Appel, K., & Haken, W. (1977). Every Planar Map is Four Colorable. *Illinois Journal of Mathematics, 21*, 429–567.

Benacerraf, P. (1967). God, the Devil, and Gödel. *The Monist, 51*, 9–32.

Boolos, G. (1993) *The logic of provability*. Cambridge University Press.

Carlson, T.J. (1999). Ordinal arithmetic and $\Sigma_1$ elementarity. *Archive for Mathematical Logic, 38*, 449–460.

Carlson, T.J. (2000). Knowledge, machines, and the consistency of Reinhardt’s strong mechanistic thesis. *Annals of Pure and Applied Logic, 105*, 51–82.

Carlson, T.J. (2001). Elementary patterns of resemblance. *Annals of Pure and Applied Logic, 108*, 19–77.

Carlson, T.J. (2012). Sound Epistemic Theories and Collapsing Knowledge. Slides from the Workshop on *The Limits and Scope of Mathematical Knowledge* at the University of Bristol.

Chow, T.Y. (1998). The surprise examination or unexpected hanging paradox. *The American Mathematical Monthly, 105*, 41–51.

Costa-Leite, A. (2006). Fusions of Modal Logics and Fitch’s Paradox. *Croatian Journal of Philosophy, 6*, 281–290.

Égré, P. (2005). The knower paradox in light of provability interpretations of modal logic. *Journal of Logic, Language, and Information, 14*, 13–48.

Fitch, F. (1963). A logical analysis of some value concepts. *The Journal of Symbolic Logic, 28*, 135–142.

Fitting, M. (2007). Modal proof theory. In Blackburn, P., van Benthem, J., & Wolter, F. (Eds.), *Handbook of modal logic*, pp. 85–138. Elsevier.

van Fraassen, B.C. (2011). Thomason’s Paradox for Belief, and Two Consequence Relations. *The Journal of Philosophical Logic, 40*, 15–32.

Friedman, H. (1975). The disjunction property implies the numerical existence property. *Proceedings of the National Academy of the Sciences, 72*, 2877–2878.

Friedman, H., & Sheard, M. (1987). An Axiomatic Approach to Self-Referential Truth. *Annals of Pure and Applied Logic, 33*, 1–21.

Gödel, K. (1951). Some basic theorems on the foundations of mathematics and their philosophical implications. 24th *Josia Willard Gibbs Lecture*.

Hakli, R., & Negri, S. (2012). Does the deduction theorem fail for modal logic? *Synthese, 187*, 849–867.

Halbach, V. (2008). On a side effect of solving Fitch’s paradox by typing knowledge. *Analysis, 68*, 114–120.

Halbach, V., & Welch, P. (2009). Necessities and necessary truths: A prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind, 118*, 71–100.

Halpern, J., & Moses, Y. (1986). Taken by surprise: The paradox of the surprise test revisited. *Journal of Philosophical Logic, 15*, 281-304.

Horsten, L., & Leitgeb, H. (2001). No Future. *Journal of Philosophical Logic, 30*, 259–265.
Kaplan, D., & Montague, R. (1960). A Paradox Regained. *Notre Dame Journal of Formal Logic*, 1, 79-90.

Kirby, L., & Paris, J. (1982). Accessible independence results for Peano arithmetic. *Bulletin of the London Mathematical Society*, 14, 285–293.

Kritchman, S., & Raz, R. (2010). The Surprise Examination Paradox and the Second Incompleteness Theorem. *Notices of the AMS*, 57, 1454-1458.

Lucas, J.R. (1961). Minds, machines, and Gödel. *Philosophy*, 36, 112–137.

McLelland, J., & Chihara, C. (1975). The surprise examination paradox. *Journal of Philosophical Logic*, 4, 71–89.

Montague, R. (1963). Syntactic treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, 16, 153–167.

Myhill, J. (1960). Some Remarks on the Notion of Proof. *The Journal of Philosophy*, 57, 461–471.

Penrose, R. (1989). *The emperor’s new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.

Putnam, H. (2006). After Gödel. *Logic Journal of the IGPL*, 14, 745–754.

Reinhardt, W. (1985a). Absolute versions of incompleteness theorems. *Nous*, 19, 317–346.

Reinhardt, W. (1985b). The consistency of a variant of Church’s thesis with an axiomatic theory of an epistemic notion. *Revista Colombiana de Matematicas*, 19, 177–200.

Salerno, J. (2010). Introduction to knowability and beyond. *Synthese*, 173, 1-8.

Shapiro, S. (1988). Epistemic and intuitionistic arithmetic. In Shapiro S. (ed), *Intensional Mathematics*, pp. 11–46. Elsevier.

Shapiro, S. (1998). Incompleteness, Mechanism, and Optimism. *The Bulletin of Symbolic Logic*, 4, 273–302.

Shapiro, S. (2012). Idealization, mechanism, and knowability. Slides from the Workshop on The Limits and Scope of Mathematical Knowledge at the University of Bristol.

Smorynski, C. (1984). Modal logic and self-reference. In Gabbay, D., & Guenthner, F. (Eds.), *Handbook of philosophical logic* Volume II, 441–495.

Thomason, R. (1980). A note on syntactical treatments of modality. *Synthese*, 44, 391–395.

Department of Mathematics
The Ohio State University
231 W. 18th Ave., Columbus, Ohio, 43210, USA

E-mail: alexander@math.ohio-state.edu