Chapter 34
Some Computational Constraints in Epistemic Logic

Timothy Williamson

Introduction

This paper concerns limits that some epistemic logics impose on the complexity of an epistemic agent's reasoning, rather than limits on the complexity of the epistemic logic itself.

As an epistemic agent, one theorizes about a world which contains the theorizing of epistemic agents, including oneself. Epistemic logicians theorize about the abstract structure of epistemic agents' theorizing. This paper concerns the comparatively simple special case of epistemic logic in which only one agent is considered. Such an epistemic agent theorizes about a world which contains that agent's theorizing. One has knowledge about one's own knowledge, or beliefs about one's own beliefs. The considerations of this paper can be generalized to multi-agent epistemic logic, but that will not be done here. Formally, single-agent epistemic logic is just standard monomodal logic; we call it 'epistemic' in view of the envisaged applications.

In epistemic logic, we typically abstract away from some practical computational limitations of all real epistemic agents. For example, we are not concerned with their failure to infer from a proposition q the disjunction $q \vee r$ for every unrelated proposition r. What matters is that if some propositions do in fact follow from the agent's theory (from what the agent knows, or believes), then so too do all their logical consequences. For ease of exposition, we may idealize epistemic agents and describe them as knowing whatever follows from what they know, or as believing whatever follows from what they believe, but we could equally well redescribe the matter in less contentious terms by substituting 'p follows from what one knows'

T. Williamson (✉)
University of Oxford, Oxford, UK
e-mail: timothy.williamson@philosophy.ox.ac.uk

H. Arló-Costa et al. (eds.), *Readings in Formal Epistemology*, Springer Graduate Texts in Philosophy 1, DOI 10.1007/978-3-319-20451-2_34

for 'one knows *p*' or '*p* follows from what one believes' for 'one believes *p*' throughout the informal renderings of formulas, at the cost only of some clumsiness. Thus, if we so wish, we can make what looks like the notorious assumption of logical omniscience true by definition of the relevant epistemic operators. On suitable readings, it is a triviality rather than an idealization. It does not follow that no computational constraints are of any concern to epistemic logic. For if one's knowledge is logically closed by definition, that makes it computationally all the harder to know that one does *not* know something: in the standard jargon, logical omniscience poses a new threat to negative introspection. That threat is one of the phenomena to be investigated in this paper.

In a recursively axiomatizable epistemic logic, logical omniscience amounts to closure under a recursively axiomatizable system of inferences. Thus all the inferences in question can in principle be carried out by a single Turing machine, an idealized computer. Epistemic logicians do not usually want to make assumptions which would require an epistemic agent to exceed every Turing machine in computational power. In particular, such a requirement would presumably defeat the purpose of the many current applications of epistemic logic in computer science. By extension, epistemic logicians might prefer not to make assumptions which would permit an epistemic agent not to exceed every Turing machine in computational power only under highly restrictive conditions. Of course, such assumptions might be perfectly appropriate in special applications of epistemic logic to cases in which those restrictive conditions may be treated as met. But they would not be appropriate in more general theoretical uses of epistemic logic.

As an example, let us consider the so-called axiom of *negative introspection* alluded to above. It may be read as the claim that if one does not know *p* then one knows that one does not know *p*, or that if one does not believe *p* then one believes that one does not believe *p*. In terms of theories: if one's theory does not entail *p*, then one's theory entails that one's theory does not entail *p*. That assumption is acceptable in special cases for special values of '*p*'. However, for a theory to be consistent is in effect for there to be some *p* which it does not entail. On this reading, negative introspection implies that if one's theory is consistent then it entails its own consistency. But, by Gödel's second incompleteness theorem, if one's theory is recursively axiomatizable and includes Peano arithmetic, then it entails its own consistency only if it is inconsistent. Thus, combined with the incompleteness theorem, negative introspection implies that if one's theory is recursively axiomatizable then it includes Peano arithmetic only if it is inconsistent. Yet, in a wide range of interesting cases, the output of a Turing machine, or the theory of an epistemic agent of equal computational power, is a consistent recursively axiomatizable theory which includes Peano arithmetic. Thus, except in special circumstances, the negative introspection axiom imposes an unwarranted constraint on the computational power of epistemic agents.

Naturally, such an argument must be made more rigorous before we can place much confidence in it. That will be done below. The problem for the negative introspection axiom turns out to be rather general: it arises not just for extensions of Peano arithmetic but for any undecidable recursively axiomatizable theory, that is, for any theory which is the output of some Turing machine while its complement is not. It is very natural to consider epistemic agents whose theories are of that kind.

The aim of this paper is not primarily to criticize the negative introspection axiom. Rather, it is to generalize the problem to which that axiom gives rise, to formulate precisely the conditions which a system of epistemic logic must satisfy in order not to be susceptible to such problems, and to investigate which systems satisfy those conditions. The conditions in question will be called *r.e. conservativeness* and *r.e. quasi-conservativeness*. Very roughly indeed, a system satisfies these conditions if it has a wide enough variety of models in which the epistemic agent is computationally constrained. Such models appear to be among the intended models on various applications of epistemic logic. As already noted, systems of epistemic logic which do not satisfy the conditions may be appropriate for other applications. But it is time to be more precise.

Elementary Epistemic Logic

Let L be the language consisting of countably many propositional variables p_0, p_1, p_2, … (p and q represent arbitrary distinct variables), the falsity constant $\perp$ and the material conditional $\supset$. Other operators are treated as metalinguistic abbreviations in the usual way. We expand L to the language $L_\Box$ of propositional modal logic by adding the operator $\Box$. $\Diamond\alpha$ abbreviates $\neg\Box\neg\alpha$. Unless otherwise specified, the metalinguistic variables α, β, γ, … range over all formulas of $L_\Box$. We use the necessity symbol $\Box$ from modal logic to make various formulas and formal systems look familiar, without prejudice to its interpretation. We reinterpret $\Box$ as something like ‘I know that’ or ‘I believe that’. To generalize over reinterpretations, we use the neutral verb ‘cognize’ for $\Box$ in informal renditions of formulas.

A *theory* in $L_\Box$ is a subset of $L_\Box$ containing all truth-functional tautologies and closed under modus ponens for $\supset$ (MP). A *model* M of $L_\Box$ induces a function $M() : L_\Box \to \{0, 1\}$ where $M(\perp) = 0$ and $M(\alpha \supset \beta) = 1$ if and only if $M(\alpha) \leq M(\beta)$. Intuitively, $M(\alpha) = 1$ if and only if α is true in M; $M(\alpha) = 0$ if and only if α is false in M. An *application* of epistemic logic determines a class of its intended models. The logic of the application is the set of formulas α such that $M(\alpha) = 1$ for every intended model M; thus the logic is a theory in $L_\Box$. Of course, we can also define a relation of logical consequence on the models, but for present purposes it is simpler to identify a logic with the set of its theorems.

Since atomic sentences are treated simply as propositional variables, we may substitute complex formulas for them. More precisely, we assume that for each intended model M and uniform substitution σ there is an intended model M^σ such that for every α $M^\sigma(\alpha) = M(\sigma\alpha)$. Thus the logic of the application is closed under uniform substitution (US).

A *modal logic* is a theory in $L_\Box$ closed under US. The logic of an application is a modal logic. The smallest modal logic is PC, the set of all truth-functional tautologies. If Σ is a modal logic, we write $\vdash_\Sigma \alpha$ when $\alpha \in \Sigma$. For any $\mathrm{X} \subseteq \mathrm{L}_\Box$, we define $\mathrm{X} \vdash_\Sigma \alpha$ if and only if $\vdash_\Sigma \wedge \mathrm{X}_0 \supset \alpha$ for some finite $\mathrm{X}_0 \subseteq \mathrm{X}$ ($\wedge \mathrm{X}_0$ and $\vee \mathrm{X}_0$ are the conjunction and disjunction respectively of X_0 on a fixed ordering of the language). X is Σ-consistent unless $\mathrm{X} \vdash_\Sigma \bot$. A maximal Σ-consistent set is a Σ-consistent set not properly included in any Σ-consistent set.

If M is a model, let $\Box^{-1}\mathrm{M} = \{\alpha : \mathrm{M}(\Box\alpha) = 1\}$. Thus $\Box^{-1}\mathrm{M}$ expresses what the agent cognizes in M. If Σ is the logic of an application on which $\Box^{-1}\mathrm{M}$ is a theory in $\mathrm{L}_\Box$ for every intended model M, then for all formulas α and β, $\vdash_\Sigma \Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ (axiom schema K) and if $\vdash_{\mathrm{PC}} \alpha$ then $\vdash_\Sigma \Box\alpha$ (rule $\mathrm{RN}_{\mathrm{PC}}$). A modal logic satisfying $\mathrm{RN}_{\mathrm{PC}}$ and K is *prenormal*. If cognizing is knowing or believing, then prenormality is an extreme idealization, a form of logical omniscience. But if cognizing is the closure of knowing or believing under at least truth-functional consequence, then prenormality is innocuous. The rule $\mathrm{RN}_{\mathrm{PC}}$ differs from the stronger and better-known rule RN (necessitation or epistemization): if $\vdash_\Sigma \alpha$ then $\vdash_\Sigma \Box\alpha$. A modal logic Σ satisfying RN and K is *normal*. Unlike $\mathrm{RN}_{\mathrm{PC}}$, RN requires the agent to cognize all theorems of the epistemic logic, not just all truth-functional tautologies. For instance, $\Box\top$ is a theorem of every prenormal logic by $\mathrm{RN}_{\mathrm{PC}}$, but since it is not a theorem of PC we cannot iterate the rule; $\Box\Box\top$ is not a theorem of the smallest prenormal logic. By contrast, we can iterate RN, and $\Box\Box\top$ is a theorem of every normal modal logic. Prenormality does not imply that agents cognize their own cognizing. It merely implies that they can formulate propositions about cognizing, for since $\Box\alpha \supset \Box\alpha$ is a truth-functional tautology, $\Box(\Box\alpha \supset \Box\alpha)$ is a theorem of every prenormal logic. Since normality entails prenormality, results about all prenormal logics apply to all normal modal logics. Every logic extending a prenormal logic is prenormal; by contrast, some nonnormal logics extend normal logics, although any extension of a normal logic is at least prenormal.

Any normal logic Σ has a possible worlds semantics where $\Box\alpha$ is true at a world w in a model M if and only if α is true at every world in M to which w has the accessibility relation of M. Intuitively, a world x is accessible from w if and only if what the agent cognizes at w is true at x. In other words, one world is accessible from another if and only if for all one cognizes in the latter one is in the former. The formulas α such that $\Box\alpha$ is true at w express what the agent cognizes at w. For every normal logic Σ there is a class C of models such that Σ consists of exactly the formulas true at every world in every model in C.

Many authors require the accessibility relation to be an equivalence relation (reflexive, symmetric and transitive) for every intended model of their application. A common attitude is expressed by the authors of a standard text, who write that the postulate 'seems reasonable for many applications we have in mind', but 'we can certainly imagine other possibilities' (Fagin et al. 1995, 33). For example, if x is accessible from w if and only if appearances to the agent are identical in x and w, then accessibility is an equivalence relation because identity in any given respect is an equivalence relation. The logic of the class of all possible worlds models in which accessibility is an equivalence relation is the modal system known as S5 : $\vdash_{\mathrm{S5}} \alpha$ if and only if α is true in every model for which accessibility is an equivalence relation. Since equivalence relations correspond to partitions of the set of worlds, S5 is also known as the logic of the *partitional* conception of knowledge. S5 is the smallest normal modal logic with the theorem schemas T($\Box\alpha \supset \alpha$) and E ($\neg\Box\alpha \supset \Box\neg\Box\alpha$). T (*truthfulness*) says that the agent cognizes only truths; it is appropriate for applications on which one cognizes only what follows from what one knows. T corresponds to the condition that accessibility be reflexive. For applications on which one cognizes what follows from what one *believes*, T would have to be dropped, perhaps replaced by the weaker principle D ($\Box\alpha \supset \Diamond\alpha$). D requires cognition to be consistent in the sense that an agent who cognizes something does not also cognize its negation. D corresponds to the condition that accessibility be serial (from every world some world is accessible). E is the principle of *negative introspection*: cognition records its omissions in the sense that agents who do not cognize something cognize that they do not cognize it. E corresponds to the condition that accessibility be euclidean (worlds accessible from a given world are accessible from each other). In S5 we can derive the principle of *positive introspection* 4 ($\Box\alpha \supset \Box\Box\alpha$), that cognition records its contents in the sense that agents who cognize something cognize that they cognize it. 4 corresponds to the condition that accessibility be transitive. If T is dropped or weakened to D then 4 is no longer derivable from E, so 4 might be added as an independent schema. Accessibility is reflexive (T) and euclidean (E) if and only if it is an equivalence relation.

Computational Constraints

To formulate computational constraints, we generalize concepts from recursion theory to $L_\Box$ using a standard intuitively computable coding procedure. A model M is r.e. if and only if $\Box^{-1}$M (which expresses what the agent cognizes in M) is an r.e. (recursively enumerable) theory in $L_\Box$. In that sense, the agent's cognition in an r.e. model does not exceed the computational capacity of a sufficiently powerful Turing machine.

Consider the restriction of $\Box^{-1}$M to the $\Box$-free sublanguage L, L $\cap$ $\Box^{-1}$M. Let $\Box^{-1}$M be an r.e. theory in $L_\Box$. Thus L $\cap$ $\Box^{-1}$M is an r.e. theory in L. It is the part of the agent's overall theory in M which is not specifically epistemic. From the standpoint of general epistemic logic, can we reasonably impose any further constraints on L $\cap$ $\Box^{-1}$M beyond recursive enumerability?

If $\Box^{-1}$M is required to be consistent, L $\cap$ $\Box^{-1}$M is consistent too. Can we limit the possible values of L $\cap$ $\Box^{-1}$M still further? For many applications we cannot. L $\cap$ $\Box^{-1}$M simply expresses what the agent cognizes in M about some aspect of reality. The agent can store any r.e. theory in L as a recursive axiomatization (Craig 1953). If the agent might cognize that aspect of reality simply by having learned a theory about it on the testimony of a teacher, any (consistent) r.e. theory in L is possible. In particular, we can interpret the propositional variables as mutually independent. For example, given a black box which may or may not flash a light on input of a symbol for a natural number, we can read p_i as 'The light flashes on input *i*'. Then any (consistent) r.e. theory in L could exhaust everything expressible in L which the agent (with only the computational power of a Turing machine) has learned about the black box. Such situations seem quite reasonable. If we want an epistemic logic to have a generality beyond some local application, it should apply to them: such situations should correspond to intended models. Now any application which has all those intended models thereby satisfies (*) or ($*_{\text{con}}$), depending on whether the epistemic agent's theory is required to be consistent:

(*) For every r.e. theory R in L, $L \cap \Box^{-1}M = R$ for some r.e. intended model M.

($*_{\text{con}}$) For every consistent r.e. theory R in L, $L \cap \Box^{-1}M = R$ for some r.e. intended model M.

(*) is appropriate for readings of □ like 'It follows from what I believe that ... ', if the agent is not required to be consistent. For readings of □ like 'It follows from what I know that ... ', only ($*_{\text{con}}$) is appropriate, for one can know only truths and any set of truths is consistent. We can define corresponding constraints on a modal logic Σ without reference to models:

Σ is *r.e. conservative* if and only if for every r.e. theory R in L, there is a maximal Σ-consistent set X such that $\Box^{-1}X$ is r.e. and $L \cap \Box^{-1}X = R$.

Σ is *r.e. quasi-conservative* if and only if for every consistent r.e. theory R in L, there is a maximal Σ-consistent set X such that $\Box^{-1}X$ is r.e. and $L \cap \Box^{-1}X = R$.

Here $\Box^{-1}X = \{\alpha \in L_{\Box} : \Box\alpha \in X\}$. Roughly, if Σ is r.e. (quasi-)conservative then every (consistent) r.e. theory in the language without □ is conservatively extended by an r.e. theory in the language with □ such that it is consistent in Σ for R to be exactly what the agent cognizes in the language without □ while what the agent cognizes in the language with □ constitutes an r.e. theory. If an application satisfies (*), its logic is r.e. conservative, for X can be the set of formulas true in M. Conversely, any r.e. conservative logic is the logic of some application which satisfies (*), for some appropriate kind of intended model. The same relationships hold between ($*_{\text{con}}$) and r.e. quasi-conservativeness. For many applications of epistemic logic, the class of intended models is quite restricted and even ($*_{\text{con}}$) does not hold. But if the application interprets □ as something like 'It follows from what I believe/know that', without special restrictions on the epistemic subject, then situations of the kind described above will correspond to intended models and the logic of the application will be r.e. [quasi-] conservative. In this paper we do not attempt to determine which informally presented applications of epistemic logic satisfy (*) or ($*_{\text{con}}$). We simply investigate which logics are r.e. [quasi-] conservative.

Trivially, every r.e. conservative modal logic is r.e. quasi-conservative. Examples will be given below of r.e. quasi-conservative normal modal logics which are not r.e. conservative. For prenormal modal logics, r.e. conservativeness can be characterized in terms of r.e. quasi-conservativeness in a simple way which allows us to transfer results about one to the other:

Proposition 34.1 Let Σ be a prenormal modal logic. Then Σ is r.e. conservative if and only if Σ is r.e. quasi-conservative and not $\vdash_{\Sigma} \Diamond\top$.

Proof Let $\Box$ L $=\{\Box\alpha : \alpha \in$ L$\}$. ($\Rightarrow$) Trivially, Σ is r.e. quasi-conservative if r.e. conservative. Suppose that $\vdash_\Sigma \Diamond\top$. Since $\Box$ L $\vdash_\Sigma \Box\neg\top$, $\Box$ L is Σ-inconsistent. Thus L $\cap\ \Box^{-1}$X $=$ L for no Σ-consistent set X. Since L is an r.e. theory in L, Σ is not r.e. conservative. ($\Leftarrow$) Suppose that Σ is r.e. quasi-conservative but not r.e. conservative. Since L is the only inconsistent theory in L, there is no maximal Σ-consistent set X such that $\Box^{-1}$X is r.e. and L $\cap\ \Box^{-1}$X $=$ L. If $\Box$ L is Σ-consistent, then some maximal Σ-consistent set X extends $\Box$ L, so L $\cap\ \Box^{-1}$X $=$ L; but for α $\in$ L $\Box \vdash_\Sigma \bot \supset \alpha$, so $\vdash_\Sigma \Box\bot \supset \Box\alpha$ by prenormality, so $\Box^{-1}$X $=$ L$_\Box$ because $\Box$ $\bot \in X$, so $\Box^{-1}$X is r.e. Thus $\Box$ L is Σ-inconsistent, i.e., for some $\alpha_0, \ldots, \alpha_m \in$ L, $\vdash_\Sigma \neg\wedge\{\Box\alpha_i : i \leq m\}$. But for $i \leq m$, $\vdash_\Sigma \Box\neg\top \supset \Box\alpha_i$ by prenormality, so $\vdash_\Sigma \neg\Box\neg\top$.

Examination of the proof shows that the prenormality condition can be weakened to this: if $\vdash_{PC}\ \alpha \supset \beta$ then $\vdash_\Sigma\ \Box\alpha \supset \Box\beta$. An example of a reading of $\Box$ which verifies this weaker condition but falsifies prenormality is ‘There is a subjective probability of at least x that’, where $0 < x < 1$, for prenormality implies that $\vdash_\Sigma(\Box p \wedge \Box q) \supset \Box(p \wedge q)$, whereas this reading invalidates that formula. Prenormality can be weakened in similar ways for subsequent propositions.

R.e. conservativeness and r.e. quasi-conservativeness do not state upper or lower bounds on the epistemic agent’s computational capacity. Rather, they state upper bounds on the strength of the epistemic logic itself; evidently a modal logic with an r.e. [quasi-] conservative extension is itself r.e. [quasi-] conservative. But too strong a logic can impose unwarranted restrictions on the agent’s theory of the world given an upper bound on the agent’s computational capacity.

Some Non-R.e. Quasi-Conservative Logics

Which modal logics are not r.e. [quasi-] conservative? Obviously, since $\vdash_{S5}\ \Diamond\top$, the logic S5 is not r.e. conservative. Since S5 is decidable, this does not result from non-recursiveness in S5 itself. More significantly:

Proposition 34.2 S5 is not r.e. quasi-conservative.

Proof (Skyrms 1978, 377 and Shin and Williamson 1994, Proposition 34.1 have similar proofs of related facts about S5): Let R be a non-recursive r.e. theory in L; R is consistent. Suppose that $\Box^{-1}$X is r.e. and L $\cap\ \Box^{-1}$X $=$ R for some maximal S5-consistent set X. Now L $-$ R $=\{\alpha : \Box\neg\Box\alpha \in$ X$\}$ $\cap$ L. For if $\alpha \in$ L $-$ R then $\Box\alpha \notin$ X, so $\neg\Box\alpha \in$ X; but $\vdash_{S5}\ \neg\Box\alpha \supset \Box\neg\Box\alpha$, so $\Box\neg\Box\alpha \in$ X since X is maximal S5-consistent. Conversely, if $\Box\neg\Box\alpha \in$ X then $\neg\Box\alpha \in$ X since $\vdash_{S5}$ $\Box\neg\Box\alpha \supset \neg\Box\alpha$, so $\Box\alpha \notin$ X, so $\alpha \notin$ R since L $\cap\ \Box^{-1}$X $=$ R. Since $\Box^{-1}$X is r.e., so is $\{\alpha : \Box\neg\Box\alpha \in$ X$)$ $\cap$ L, i.e. L – R. Contradiction.

Thus the partitional conception of knowledge prevents a subject with the computational capacity of a Turing machine from having as the restriction of its theory to the $\Box$-free language any non-recursive r.e. theory (for other problems with the S5 schema in epistemic logic and further references see Williamson (2000, 23–24, 166–167, 226–228, 316–317)). Thus S5 is unsuitable as a general epistemic logic for Turing machines.

The proof of Proposition 34.2 depends on the existence of an r.e. set whose complement is not r.e. By contrast, the complement of any recursive set is itself recursive; decidability, unlike semi-decidability, is symmetric between positive and negative answers. The analogue of Proposition 34.2 for a notion like r.e. quasi-conservativeness but defined in terms of recursiveness rather than recursive enumerability would be false. For it is not hard to show that if R is a consistent recursive theory in L, then there is a maximal S5-consistent set X in $\mathrm{L}_\Box$ such that $\Box^{-1}\mathrm{X}$ is recursive and $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$. Thus S5 imposes computational constraints not on very clever agents (whose theories need not be r.e.) or on very stupid agents (whose theories must be recursive) but on half-clever agents (whose theories must be r.e. but need not be recursive).

Proposition 34.2 is the rigorous version of the argument sketched in the introduction. Can we generalize it? The next result provides a rather unintuitive necessary condition for r.e. quasi-conservativeness which nevertheless has many applications.

Theorem 34.3 Let Σ be a modal logic such that for some formulas $\alpha_0, \ldots, \alpha_n \in \mathrm{L}_\Box$ and $\beta_0, \ldots, \beta_n \in \mathrm{L}$, $\vdash_\Sigma \vee\{\Box\alpha_i : i \leq n\}$ and, for each $i \leq n$, $\vdash_\Sigma (\Box\alpha_i \wedge \Box\beta_i) \supset \Box\bot$ and not $\vdash_{\mathrm{PC}} \neg\beta_i$. Then Σ is not r.e. quasi-conservative.

Proof There are pairwise disjoint r.e. subsets $\mathrm{I}_0, \mathrm{I}_1, \mathrm{I}_2, \ldots$ of the natural numbers $\mathbb{N}$ such that for every total recursive function f, $i \in \mathrm{I}_{f(i)}$ for some $i \in \mathbb{N}$. For let $f[0]$, $f[1], f[2], \ldots$ be a recursive enumeration of all partial and total recursive functions on $\mathbb{N}$ and set $\mathrm{I}_i = \{j : f[j](j)$ is defined and $= i\}$; then $j \in \mathrm{I}_{f[j](i)}$ whenever $f[j]$ is total, I_i is r.e. and $\mathrm{I}_i \cap \mathrm{I}_j = \{\}$ whenever $i \neq j$. Now suppose that (i) $\vdash_\Sigma \vee\{\Box\alpha_i : i \leq n\}$; (ii) $\vdash_\Sigma (\Box\alpha_i \wedge \Box\beta_i) \supset \Box\bot$ for each $i \leq n$; (iii) $\vdash_{\mathrm{PC}} \neg\beta_i$ for no $i \leq n$. Let m be the highest subscript on any propositional variable occurring in $\beta_0, \ldots, \beta_n$. For all $i \in \mathbb{N}$, let σ_i and τ_i be substitutions such that $\sigma_i\, p_j = p_{i(m+1)+j}$ and $\tau_i\, p_{i(m+1)+j} = p_j$ for all $j \in \mathbb{N}$. Set $\mathrm{U} = \{\sigma_i\beta_j : i \in \mathrm{I}_j\}$. Since the σ_i are recursive and the I_j are r.e., U is r.e. Now $\vdash_{\mathrm{PC}} \neg\sigma_i\beta_j$ for no i, j, otherwise $\vdash_{\mathrm{PC}} \neg\tau_i\sigma_i\beta_j$, i.e., $\vdash_{\mathrm{PC}} \neg\beta_j$, contrary to (iii). Moreover, if $h \neq i$ then $\sigma_h\beta_j$ and $\sigma_i\beta_k$ have no propositional variable in common. Thus if $h \in \mathrm{I}_j$ and $i \in \mathrm{I}_k$ and $\sigma_h\beta_j$ has a variable in common with $\sigma_i\beta_k$, then $h = i$, so $j = k$ because the I_j are pairwise disjoint. Hence no two members of U have a propositional variable in common. Thus U is consistent. Let R be the smallest theory in L containing U; R is consistent and r.e. Suppose that for some maximal Σ-consistent set X, $\Box^{-1}\mathrm{X}$ is r.e. and $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$. Let the total recursive function g enumerate $\Box^{-1}\mathrm{X}$. Fix $j \in \mathbb{N}$. By (i), $\vdash_\Sigma \vee\{\Box\sigma_j\alpha_i : i \leq n.\}$ since Σ is closed under US, so $\Box\sigma_j\alpha_i \in \mathrm{Y}$ for some $i \leq n$ since Y is maximal Σ-consistent. Thus $g(k) = \sigma_j\alpha_i$ for some k; let $k(j)$ be the least k such that $g(k) \in \{\sigma_j\alpha_i : i \leq n\}$. Let $f(j)$ be the least $i \leq n$ such that $g(k(j)) = \sigma_j\alpha_i$. Since g enumerates $\Box^{-1}\mathrm{X}$, $\Box\sigma_j\alpha_{f(j)} \in \mathrm{X}$. Since g and σ_j are total recursive, k is total recursive, so f is total recursive. Thus $j \in \mathrm{I}_{f(j)}$ for some $j \in \mathbb{N}$, so $\sigma_j\beta_{f(j)} \in \mathrm{U} \subseteq \mathrm{R}$ since $f(j) \leq n$. Since $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$, $\Box\sigma_j\beta_{f(j)} \in \mathrm{X}$. By (ii), $\vdash_\Sigma (\Box\alpha_{f(j)} \wedge \Box\beta_{f(j)}) \supset \Box\bot$, so $\vdash_\Sigma (\Box\sigma_j\alpha_{f(j)} \wedge \Box\sigma_j\beta_{f(j)}) \supset \Box\bot$; since X is maximal Σ-consistent, $\Box\bot \in \mathrm{X}$. Thus $\bot \in \mathrm{R}$, contradicting the consistency of R. Thus no such set as X can exist, so Σ is not r.e. quasi-conservative.

In other words, a necessary condition for Σ to be r.e. quasi-conservative is that for all formulas $\alpha_0, \ldots, a_n \in L_\Box$ and $\beta_0, \ldots, \beta_n \in L$, if $\vdash_\Sigma \vee(\Box\alpha_j : i \leq n)$ and, for each $i \leq n$, $\vdash_\Sigma(\Box\alpha_i \wedge \Box\beta_i) \supset \Box\bot$ then, for some $i \leq n$, $\vdash_{PC} \neg\beta_i$. Of course, if Σ is prenormal and contains the D axiom (requiring the agent to be consistent) then the condition that $\vdash_\Sigma(\Box\alpha_i \wedge \Box\beta_i) \supset \Box\bot$ can be simplified to the condition that $\vdash_\Sigma \neg\Box(\alpha_i \wedge \beta_i)$.

Open Problem Is the necessary condition for r.e. quasi-conservativeness in Theorem 34.3 (or some natural generalization of it) also sufficient?

Observation The proof of Theorem 34.3 uses significantly more recursion theory than does the proof of Proposition 34.2, which relies only on the existence of an r.e. set whose complement is not r.e. Samson Abramsky observed (informal communication) that the proof of Proposition 34.2 would generalize to a setting in which r.e. sets were replaced by open sets in a suitable topology (in which not all open sets have open complements). It would be interesting to see whether a generalization along such lines yielded a smoother theory. One might then seek an intuitive interpretation of the topology.

To see that Proposition 34.2 is a special case of Theorem 34.3, put $n = 1$, $\alpha_0 = \Diamond\neg p$, $\alpha_1 = \Diamond p$, $\beta_0 = p$ and $\beta_1 = \neg p$. Now; $\vdash_{S5} \Box\Diamond\neg p \vee \Box\Box p$; $\vdash_{S5} (\Box\Diamond\neg p \wedge \Box p) \supset \Box\bot$ because $\vdash_{S5} \Box p \supset \Box\Box p$ and S5 is normal; likewise $\vdash_{S5} (\Box\Diamond p \wedge \Box\neg p) \supset \Box\bot$; finally, neither $\vdash_{S5} p$ nor $\vdash_{S5} \neg p$. These features of S5 follow easily from the fact that it is a consistent normal extension of $K4G_1$, the smallest normal logic Σ including both 4 and G_1 $(\Diamond\Box\alpha \supset \Diamond\Box\alpha)$. Since the inconsistent logic is certainly not r.e. quasi-conservative, we have this generalization of Proposition 34.2:

Corollary 34.4 No normal extension of KG_14 is r.e. quasi-conservative.

We can use Corollary 34.1 to show several familiar weakenings of S5 not to be r.e. quasi-conservative. G_1 corresponds to the condition that accessibility be convergent, in the sense that if x and y are both accessible from w, then some world z is accessible from both x and y. Informally, G_1 says that agents either cognize that they do not cognize α or cognize that they do not cognize $\neg\alpha$. Any normal logic satisfying E also satisfies G_1, so Corollary 34.4 implies in particular the failure of r.e. quasi-conservativeness for the logics K4E and KD4E. Those two logics are the natural analogues for belief of S5 as a logic for knowledge, since they retain positive and negative introspection while dropping truthfulness altogether (K4E) or weakening it to consistency (KD4E). Thus they are often used as logics of belief. But positive and negative introspection together violate the computational constraint in a normal logic even in the absence of truthfulness. Thus, in a generalized context, K4E or KD4E impose unacceptably strong computational constraints as logics of belief, just as S5 does as a logic of knowledge.

For more examples, consider the schemas

$$B(\alpha \supset \Box\Diamond\alpha) \text{ and } D_1(\Box(\Box\alpha \supset \beta) \vee \Box(\Box\beta \supset \alpha)).$$

B corresponds to the condition that accessibility be symmetric, D_1 to the condition that accessibility be connected, in the sense that if x and y are both accessible from w, then either x is accessible from y or y is accessible from x. Any normal logic satisfying B or D_1 also satisfies G_1, so KB4 and KD_14 are not r.e. quasi-conservative. *A fortiori*, the same holds if one requires the agent to be consistent or truthful by adding D or T respectively. Thus KD4E, KDG14, KTG_14 (= S4.2), KDD_14 and KTD_14 (= S4.3) are also not r.e. quasi-conservative. All these are sublogics of S5; we shall need to weaken S5 considerably to find an r.e. quasi-conservative logic.

Theorem 34.3 is also applicable to logics without positive introspection. We can use T rather than 4 to derive $(\Box\Diamond\neg p \wedge \Box p) \supset \Box\bot$, so:

Corollary 34.5 No normal extension of KTG_1 is r.e. quasi-conservative.

Again, consider Alt_n $(\vee\{\Box(\wedge\{p_j : j<i\}\supset p_i) : i \leq n\})$, e.g., Alt_2 is $\Box p_0 \vee \Box(p_0 \supset p_1) \vee \Box((p_0 \wedge p_1) \supset p_2)$. Alt_n corresponds to the condition that from each world at most n worlds be accessible; informally, the agent rules out all but n specific possibilities. Setting $\alpha_i = \wedge\{p_j : j<i\} \supset p_i$ and $\beta_i = \neg\alpha_i$ in Theorem 34.3 gives:

Corollary 34.6 For any n, no r.e. quasi-conservative prenormal modal logic contains Alt_n.

An epistemic logic which imposes an upper bound on how many possibilities the agent can countenance thereby excludes the agent from having some consistent r.e. theories about the black box.

Some R.e. Conservative Logics

Since every modal logic with an r.e. [quasi-] conservative extension is itself r.e. [quasi-] conservative, an efficient strategy is to seek very strong r.e. [quasi-] conservative logics, even if they are implausibly strong for most epistemic applications, because we can note that the weaker and perhaps more plausible logics which they extend will also be r.e. [quasi-] conservative.

A large class of r.e. conservative logics arises as follows. Let Σ be any epistemic logic. The agent might cognize each theorem of Σ. Moreover, an epistemic logic Σ^* may imply this, in that $\vdash_{\Sigma^*} \Box\alpha$ whenever $\vdash_{\Sigma} \alpha$. Σ and Σ^* may be distinct, even incompatible. For example, let Ver be the smallest normal modal logic containing $\Box\bot$. Interpreted epistemically, Ver implies that the agent is inconsistent; but Ver itself is consistent. An epistemic theory consisting just of Ver falsely but consistently self-attributes inconsistency, and an epistemic logic may report that the agent self-attributes inconsistency without itself attributing inconsistency to the agent. Thus Ver* may contain $\Box\Box\bot$ without $\Box\bot$. Similarly, let Triv be the smallest normal modal logic containing all theorems of the form $\alpha \equiv \Box\alpha$. Interpreted epistemically, Triv implies that the agent cognizes that his beliefs contain all and only truths; but Triv itself does not contain all and only truths (neither $\vdash_{\mathrm{Triv}} p$ nor $\vdash_{\mathrm{Triv}} \neg p$). Thus

Triv* may contain $\Box(p \equiv \Box p)$ without $p \equiv \Box p$. To be more precise, for any modal logics Λ and Σ let $\Lambda\Box\Sigma$ be the smallest normal extension of Λ containing $\{\Box\alpha : \vdash_{\Sigma} \alpha\}$. We will prove that if Σ is consistent and normal then $\mathrm{K}\Box\Sigma$ is r.e. conservative. $\mathrm{K}\Box\Sigma$ is an epistemic logic for theorizing about theories that incorporate the epistemic logic Σ. R.e. conservativeness implies no constraint on what epistemic logic the agent uses beyond consistency (if Σ is inconsistent, then $\mathrm{K}\Box\Sigma$ contains Alt_0 and so is not even r.e. quasi-conservative). In particular, the smallest normal logic K itself is r.e. conservative. Moreover, if Σ is consistent and normal, then $\mathrm{K4}\Box\Sigma$ is r.e. conservative; that is, we can add positive introspection. In particular, K4 itself is r.e. conservative. We prove this by proving that K□Ver and K□Triv are r.e. conservative. Since K□Ver and K□Triv contain $\Box\Box\bot$ and $\Box(p \equiv \Box p)$ respectively, they are too strong to be useful epistemic logics themselves, but equally they are strong enough to contain many other logics of epistemic interest, all of which must also be r.e. conservative. By contrast, Ver and Triv are not themselves even r.e. quasi-conservative, for $\vdash_{\mathrm{Ver}} \mathrm{Alt}_0$ and $\vdash_{\mathrm{Triv}} \mathrm{Alt}_1$.

For future reference, call a mapping ϕ from $\mathrm{L}_\Box$ into $\mathrm{L}_\Box$ *respectful* if and only if $\phi p = p$ for all propositional variables p, $\phi\bot = \bot$ and $\phi(\alpha \supset \beta) = \phi\alpha \supset \phi\beta$ for all formulas α and β.

Lemma 34.7 K□Triv is r.e. conservative.

Proof Let R be an r.e. theory in L. Let δ and κ be respectful mappings from $\mathrm{L}_\Box$ to L such that $\delta\Box\alpha = \delta\alpha$; $\kappa\Box\alpha = \top$ if $\mathrm{R} \vdash_{\mathrm{PC}} \delta\alpha$ and $\kappa\Box\alpha = \bot$ otherwise for all formulas α. (i) Axiomatize Triv with all truth-functional tautologies and formulas of the form $\alpha \equiv \Box\alpha$ as the axioms and MP as the only rule of inference (schema K and rule RN are easily derivable). By an easy induction on the length of proofs, $\vdash_{\mathrm{Triv}} \alpha$ only if $\vdash_{\mathrm{PC}} \delta\alpha$. (ii) Axiomatize K□Triv with all truth-functional tautologies and formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ and $\Box\gamma$ whenever $\vdash_{\mathrm{Triv}} \gamma$ as the axioms and MP as the only rule of inference (RN is a derived rule; its conclusion is always an axiom because the logic so defined is a sublogic of Triv). We show by induction on the length of proofs that $\vdash_{\mathrm{K\Box Triv}} \alpha$ only if $\vdash_{\mathrm{PC}} \kappa\alpha$. Basis: If $\vdash_{\mathrm{PC}} \alpha$, $\vdash_{\mathrm{PC}} \kappa\alpha$. If $\kappa\Box(\alpha \supset \beta) = \top$ and $\kappa\Box\alpha = \top$ then $\mathrm{R} \vdash_{\mathrm{PC}} \delta\alpha \supset \delta\beta$ and $\mathrm{R} \vdash_{\mathrm{PC}} \delta\alpha$, so $\mathrm{R} \vdash_{\mathrm{PC}} \delta\beta$, so $\kappa\Box\beta = \top$, so $\mathrm{R} \vdash_{\mathrm{PC}} \kappa(\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta))$; otherwise $\kappa\Box(\alpha \supset \beta) = \bot$ or $\kappa\Box\alpha = \bot$ and again $\mathrm{R} \vdash_{\mathrm{PC}} \kappa(\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta))$. If $\vdash_{\mathrm{Triv}} \gamma$ then $\vdash_{\mathrm{PC}} \delta\gamma$ by (i), so $\mathrm{R} \vdash_{\mathrm{PC}} \delta\gamma$, so $\kappa\Box\gamma = \top$, so $\vdash_{\mathrm{PC}} \kappa\Box\gamma$. Induction step: trivial. (iii) Put $Y = \{\Box\alpha \in \mathrm{L}_\Box : \mathrm{R} \vdash_{\mathrm{PC}} \delta\gamma\} \cup \{\neg\Box\alpha \in \mathrm{L}_\Box : \text{not } \mathrm{R} \vdash_{\mathrm{PC}} \delta\alpha\}$. Y is K□Triv-consistent, for if $\mathrm{Y}_0 \subseteq \mathrm{Y}$ is finite and $\vdash_{\mathrm{K\Box Triv}} \wedge \mathrm{Y}_0 \supset \bot$ then $\vdash_{\mathrm{PC}} \kappa\ (\wedge Y_0 \supset \top)$ by (ii), i.e. $\vdash_{\mathrm{PC}} \wedge\{\kappa\alpha : \alpha \in \mathrm{Y}_0\} \supset \bot$, which is impossible since $\{\kappa\alpha : \alpha \in \mathrm{Y}\} \subseteq \{\top, \neg \bot\}$. Let X be a maximal K□Triv-consistent extension of Y. By definition of Y, $\Box^{-1}\mathrm{X} = \{\alpha : \mathrm{R} \vdash_{\mathrm{PC}} \delta\alpha\}$, which is r.e. because R is r.e. and δ is recursive (although κ need not be). If $\alpha \in \mathrm{L}$, $\delta\alpha = \alpha$, so $\Box\alpha \in \mathrm{X}$ if and only if $\mathrm{R} \vdash_{\mathrm{PC}} \alpha$, i.e., if and only if $\alpha \in \mathrm{R}$ because R is a theory; thus $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$. Hence K□Triv is r.e. conservative.

Lemma 34.8 K□Ver is r.e. conservative.

Proof Like Lemma 34.7, but in place of δ use a respectful mapping λ such that $\lambda\Box\alpha = \top$.

A notable sublogic of K□Ver is GL, the smallest normal modal logic including $\Box(\Box\alpha \supset \alpha) \supset \Box\alpha$. Thus a corollary of Lemma 34.8 is that GL is r.e. conservative. GL is in a precise sense the logic of what is provable in Peano arithmetic (PA) about provability in PA (Boolos 1993 has exposition and references). More generally, if R is an ω-consistent r.e. extension of PA, then GL is the logic of what is provable in R about provability in R. Since a Turing machine's theory of arithmetic is presumably at best an ω-consistent r.e. extension of PA, GL is therefore a salient epistemic logic for Turing machines, and its r.e. conservativeness is not surprising.

Caution We must be careful in our informal renderings of results about provability logic. A provability operator creates an intensional context within which the substitution of coextensive but not provably coextensive descriptions can alter the truth-value of the whole sentence; this point applies in particular to descriptions of agents or their theories. On a provability interpretation of □, occurrences of □ within the scope of other occurrences of □ in effect involve just such occurrences of descriptions of agents or their theories in an intensional context, so which logic is validated can depend on the manner in which a given agent or theory is described. The validity of GL as an epistemic logic is relative to a special kind of descriptive self-presentation of the theory T in the interpretation of □, by a coding of its axioms and rules of inference. GL is not valid relative to some extensionally equivalent but intensionally distinct interpretations of □, e.g. the indexical reading 'I can prove that' as uttered by an epistemic subject with the computational capacity of a Turing machine (Shin and Williamson 1994; Williamson 1996, 1998).

Proposition 34.9 If Σ is a consistent normal modal logic, K□Σ and K4□Σ are r.e. conservative.

Proof By Makinson (1971), either $\Sigma \subseteq$ Triv or $\Sigma \subseteq$ Ver. Hence either K□Σ ⊆ K□Triv or K□Σ ⊆ K□Ver. But Schema 4 is easily derivable in both K□Triv and K□Ver, so K4□Σ ⊆ K□Triv or K4□Σ ⊆ K□Ver. By Lemmas 34.7 and 34.8, K□Triv and K□Ver are r.e. conservative, so K4□Σ is.

All the logics salient in this paper are decidable, and therefore r.e., but we should note that an epistemic logic need not be r.e. to be r.e. conservative:

Corollary 34.10 Not all r.e. conservative normal modal logics are r.e.

Proof (i) We show that for any normal modal logic Σ, $\vdash_\Sigma \alpha$ if and only if $\vdash_{\mathrm{K}\Box\Sigma.} \Box\alpha$. Only the $\Leftarrow$ direction needs proving. Axiomatize K□Σ with all truth-functional tautologies and formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ and $\Box\gamma$ whenever $\vdash_\Sigma \gamma$ as the axioms and MP as the only rule of inference (RN is a derived rule; its conclusion is always an axiom because the logic so defined is a sublogic of Σ). Let η be a respectful mapping from $\mathrm{L}_\Box$ to $\mathrm{L}_\Box$ such that $\eta\Box\alpha = \alpha$ for all formulas α (η is distinct from δ in the proof of Lemma 34.7 since $\eta\Box\Box p = \Box p$ whereas $\delta\Box\Box p = p$). By induction on the length of proofs, $\vdash_{\mathrm{K}\Box\Sigma} \alpha$ only if $\vdash_\Sigma \eta\alpha$. Hence $\vdash_{\mathrm{K}\Box\Sigma} \Box\alpha$ only if $\vdash_\Sigma \alpha$. (ii) By (i), for any normal modal logics Σ_1 and Σ_2, $\mathrm{K}\Box\Sigma_1 = \mathrm{K}\Box\Sigma_2$ if and only if $\Sigma_1 = \Sigma_2$. But there are continuum many consistent normal modal logics (Blok 1980 has much more on these lines). Hence there are continuum many corresponding logics of the form $K\Box\Sigma$; all are r.e. conservative by Proposition 34.9. Since only countably many modal logics are r.e., some of them are not r.e.

One limitation of Proposition 34.9 is that $K\Box\Sigma$ and $K4\Box\Sigma$ never contain the consistency schema D. In a sense this limitation is easily repaired. For any modal logic Σ, let $\Sigma\ [D]$ be the smallest extension of Σ containing D; thus $\vdash_{\Sigma[\mathrm{D}]} \alpha$ just in case $\vdash_{\Sigma} \Diamond\top \supset \alpha$.

Proposition 34.11 For any r.e. conservative modal logic Σ, Σ[D] is r.e. quasi-conservative.

Proof For any consistent theory R, any maximal Σ-consistent set X such that $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$ is Σ[D]-consistent because $\Diamond\top \in \mathrm{X}$.

Corollary 34.12 If Σ is a consistent normal modal logic, $(K\Box\Sigma)$[D] and $(K4\Box\Sigma)$[D] are r.e. quasi-conservative.

Proof By Propositions 34.9 and 34.11.

Although Σ[D] is always prenormal, it may not be normal, even if Σ is normal; sometimes not $\vdash_{\Sigma[\mathrm{D}]} \Box\Diamond\top$. But we can also consider epistemic interpretations of normal logics with the D schema, e.g., KD and KD4. Such logics contain $\Box\Diamond\top$; they require agents to cognize their own consistency. By Gödel's second incompleteness theorem, this condition cannot be met relative to a Gödelian manner of representing the theory in itself; no consistent normal extension of the provability logic GL contains D. But $\Box\Diamond\top$ is true on other epistemic interpretations; for example, we know that our knowledge (as opposed to our beliefs) does not imply a contradiction. Since $\mathrm{GL} \subseteq K\Box\mathrm{Ver}$, Proposition 34.9 does not generalize to the r.e. quasi-conservativeness of $KD\Box\Sigma$. But we can generalize Lemma 34.7 thus:

Proposition 34.13 If $\Sigma \subseteq \mathrm{Triv}$ then $KD\Box\Sigma$ and $KD4\Box\Sigma$ are r.e. quasi-conservative.

Proof It suffices to prove that $KD\Box\mathrm{Triv}$ ($=KD4\Box\mathrm{Triv}$) is r.e. quasi-conservative. Argue as for Lemma 34.1, adding $\Diamond\top$ as an axiom for $KD\Box\mathrm{Triv}$ and noting that if R is consistent then $\kappa\Box\neg\top = \bot$, so $\vdash_{\mathrm{PC}} \kappa\Diamond\top$.

In particular, KD and KD4 are themselves r.e. quasi-conservative; they are our first examples of r.e. quasi-conservative logics which are not r.e. conservative.

We now return to systems with the T schema. Since T implies D, only r.e. quasi-conservativeness is at issue. That constraint was motivated by the idea that any consistent r.e. theory in the non-modal language might be exactly the restriction of the agent's total r.e. theory to the non-modal language. On many epistemic interpretations, it is in the spirit of this idea that the agent's total theory might be true in the envisaged situation (for example, the agent's theory about the black box might be true, having been derived from a reliable witness). To require an epistemic logic Σ to leave open these possibilities is to require that Σ[T] be r.e. quasi-conservative, where Σ[T] is the smallest extension of Σ containing all instances of T. As with Σ[D], Σ[T] need not be normal even when Σ is; sometimes not

$\vdash_{\Sigma[T]} \Box(\Box\alpha \supset \alpha)$ (Williamson 1998, 113–116 discusses logics of the form Σ[T]). Agents may not cognize that they cognize only truths. Nevertheless, particularly when □ is interpreted in terms of knowledge, one might want an epistemic logic such as KT containing $\Box(\Box\alpha \supset \alpha)$.

Proposition 34.11 and Corollary 34.12 have no analogues for T in place of D. For any modal logic Σ, if $\vdash_\Sigma \alpha$ then $\vdash_{(K\Box\Sigma)[T]} \Box\alpha$, but $\vdash_{(K\Box\Sigma)[T]} \Box\alpha \supset \alpha$, so $\vdash_{(K\Box\Sigma)[T]} \alpha$; thus (K□Σ)[T] extends Σ and is r.e. quasi-conservative only if Σ is. Similarly, Proposition 34.12 would be false with T in place of D (counterexample: Σ = S5). Therefore, needing a different approach, we start with the system GL[T]. GL[T] has intrinsic interest, for it is the provability logic GLS introduced by Solovay and shown by him to be the logic of what is true (rather than provable) about provability in PA; more generally, it is the logic of what is true about provability in an ω-consistent r.e. extension of PA. GLS is therefore a salient epistemic logic for Turing machines, and its r.e. quasi-conservativeness is not surprising. Although GLS is not normal and has no consistent normal extension, we can use its r.e. quasi-conservativeness to establish that of normal logics containing T.

Proposition 34.14 GLS is r.e. quasi-conservative.

Proof Let R be an consistent r.e. theory in L. Axiomatize a theory R+ in $L_\Box$ with all members of R, truth-functional tautologies and formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$ and $\Box(\Box\alpha \supset \alpha) \supset \Box\alpha$ as the axioms and MP and RN as the rules of inference. Since R is r.e., so is R+. Let λ be the respectful mapping such that $\lambda\Box\alpha = \top$ for all formulas α. By an easy induction on the length of proofs, if $\vdash_{R+} \alpha$ then $R \vdash_{PC} \lambda\alpha$. But if $\alpha \in L$ then $\lambda\alpha = \alpha$, so $\vdash_{R+} \alpha$ only if $R \vdash_{PC} \alpha$, i.e., $\alpha \in R$; conversely, if $\alpha \in R$ then $\vdash_{R+} \alpha$; thus $L \cap R+ = R$. Let $Y \subseteq L$ be a maximal consistent extension of R. Define a set $X \subseteq L_\Box$ inductively: $p_i \in X \iff p_i \in Y$; $\bot \notin X$; $\alpha \supset \beta \in X \iff \alpha \notin X$ or $\beta \in X$; $\Box\alpha \in X \iff \vdash_{R+} \alpha$. For $\alpha \in L_\Box$, either $\alpha \in X$ or $\neg\, \alpha \in X$. We show by induction on the length of proofs that if $\vdash_{R+} \alpha$ then $\alpha \in X$. Basis: If $\alpha \in R$ then $\alpha \in Y \subseteq X$. If $\Box(\alpha \supset \beta) \in X$ and $\Box\alpha \in X$ then $\vdash_{R+} \alpha \supset \beta$ and $\vdash_{R+} \alpha$, so $\vdash_{R+} \beta$, so $\Box\beta \in X$; thus $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta) \in X$. If $\Box(\Box\alpha \supset \alpha) \in X$ then $\vdash_{R+} \Box\alpha \supset \alpha$, so $\vdash_{R+} \Box(\Box\alpha \supset \alpha)$ because R+ is closed under RN; but $\vdash_{R+} \Box(\Box\alpha \supset \alpha) \supset \Box\alpha$, so $\vdash_{R+} \Box\alpha$, so $\vdash_{R+} \alpha$, so $\Box\alpha \in X$; thus $\Box(\Box\alpha \supset \alpha) \supset \Box\alpha \in X$. Induction step: Trivial. Now axiomatize GLS with all theorems of GL and formulas of the form $\Box\alpha \supset \alpha$ as the axioms and MP as the only rule of inference. We show by induction on the length of proofs that, for all formulas α, if $\vdash_{GLS} \alpha$ then $\alpha \in X$. Basis: If $\vdash_{GL} \alpha$ then $\vdash_{R+} \alpha$ because $GL \subseteq R+$, so $\alpha \in X$ by the previous induction. If $\Box\alpha \in X$ then $\vdash_{R+} \alpha$, so again $\alpha \in X$; thus $\Box\alpha \supset \alpha \in X$. Induction step: Trivial. Hence $GLS \subseteq X$, so X is maximal GLS-consistent. Now $L \cap \Box^{-1}X = L \cap R+ = R$ and $\Box^{-1}X = R+$ is r.e. Thus GLS is r.e. quasi-conservative.

We can extend Proposition 34.14 to another system of interest in relation to provability logic. Grz is the smallest normal modal logic containing all formulas of the form $\Box(\Box(\alpha \supset \Box\alpha) \supset \alpha) \supset \alpha$. Grz turns out to be in a precise sense the logic of what is both provable and true in PA (Boolos 1993, 155–161 has all the facts about Grz used here). Grz is intimately related to GLS in a way which allows us to extend the r.e. quasi-conservativeness of GLS to Grz:

Proposition 34.15 Grz is r.e. quasi-conservative.

Proof Let R be a consistent theory in L. By Proposition 34.14, for some maximal GLS-consistent X, $L\cap\Box^{-1}X = R$ and $\Box^{-1}X$ is r.e. Let τ be the respectful mapping from $L_\Box$ to $L_\Box$ such that $\tau\Box\alpha = \Box\tau\alpha \wedge \tau\alpha$ for all formulas α. Put $\tau^{-1}X = \{\alpha : \tau\alpha \in X\}$. Now $\mathrm{Grz} \subseteq \tau^{-1}X$, for $\vdash_{\mathrm{Grz}} \alpha$ if and only if $\vdash_{\mathrm{GLS}} \tau\alpha$ (Boolos 1993, 156), so $\tau\alpha \in X$ since X is maximal GLS-consistent, so $\alpha \in \tau^{-1}X$. Since X is maximal GLS-consistent, $\tau^{-1}X$ is maximal Grz-consistent. Suppose $\alpha \in L$, so $\tau\alpha = \alpha$, so $\tau\Box\alpha = \Box\alpha \wedge \alpha$; if $\Box\alpha \in X$ then $\alpha \in X$ because $\vdash_{\mathrm{GLS}} \Box\alpha \supset \alpha$, so $\tau\Box\alpha \in X$, so $\Box\alpha \in \tau^{-1}X$; conversely, if $\Box\alpha \in \tau^{-1}X$ then $\tau\Box\alpha \in X$, so $\Box\alpha \in X$. Thus $L \cap \Box^{-1}\ \tau^{-1}X = L \cap \Box^{-1}\ X = R$. Moreover, $\Box^{-1}\ \tau^{-1}X$ is r.e. because X is r.e. and τ is recursive. Thus Grz is r.e. quasi-conservative.

Grz is not plausible as the logic of other epistemic applications. It is not a sublogic of S5 and $\vdash_{\mathrm{Grz}} \neg\Box(\Box p \wedge \Diamond\neg p)$, which in effect forbids agents to cognize that they do not cognize whether p is true. Yet you can know that what you know neither entails that the coin came down heads nor entails that it did not. However, since Grz extends the epistemically more plausible S4, the smallest normal modal logic including both the T and 4 schemas, its r.e. quasi-conservativeness entails that of S4. Truthfulness and positive introspection are together consistent with r.e. quasi-conservativeness.

Corollary 34.16 [Compare Shin and Williamson 1994 Proposition 4.] S4 is r.e. quasi-conservative.

Since S4 is r.e. quasi-conservative while S5, its extension by E, is not, and K4 is r.e. conservative while K4E is not, one might be tempted to blame E for the failure to satisfy the constraints, and to suppose that no normal logics with E is r.e. quasi-conservative. That would be a mistake; the next two propositions show that E is harmless when not combined with 4.

Proposition 34.17 KDE is r.e. quasi-conservative.

Proof Let R be a consistent r.e. theory in L. Let μ and θ be respectful mappings from $L_\Box$ to L such that for all formulas α, $\theta\Box\alpha = \top$ if $\vdash_{\mathrm{PC}} \theta\alpha$ and $\theta\Box\alpha = \bot$ otherwise; $\mu\Box\alpha = \top$ if $R \vdash_{\mathrm{PC}} \theta\alpha$ and $\mu\Box\alpha = \bot$ otherwise. Axiomatize KDE with all truth-functional tautologies and formulas of the forms $\Box(\alpha \supset \beta) \supset (\Box\alpha \supset \Box\beta)$, $\neg\Box\bot$ and $\neg\Box\alpha \supset \Box\neg\Box\alpha$ as the axioms and MP and RN as the rules of inference. We show by induction on the length of proofs that for all formulas α, $\vdash_{\mathrm{KDE}}\ \alpha$ only if $\vdash_{\mathrm{PC}}\ \theta\alpha$ and $\vdash_{\mathrm{PC}}\ \mu\alpha$. Basis: If $R \vdash_{\mathrm{PC}}\ \theta\alpha$, then $\mu\Box\alpha = \top$, so $\mu(\neg\Box\alpha \supset \Box\neg\Box\alpha) = \neg\top \supset \mu\Box\neg\Box\alpha$; if not, then not $\vdash_{\mathrm{PC}}\ \theta\alpha$, so $\theta\Box\alpha = \bot$, so $\theta\Box\alpha = \neg\bot$, so $R \vdash_{\mathrm{PC}}\ \theta\ \neg\Box\alpha$, so $\mu\Box\neg\Box\alpha = \top$, so $\mu(\neg\Box\alpha \supset \Box\neg\Box\alpha) = \neg\bot \supset \top$; either way, $\vdash_{\mathrm{PC}}\ \mu(\neg\Box\alpha \supset \Box\neg\Box\alpha)$. The rest of the induction is by now routine. The rest of the proof is like that of Lemma 34.7, with θ and μ in place of δ and κ respectively.

Corollary 34.18 KE is r.e. conservative.

Proof KE is r.e. quasi-conservative by Proposition 34.17. Since not $\vdash_{\mathrm{KE}}\ \Diamond\top$, KE is r.e. conservative by Proposition 34.1.

Although both positive and negative introspection are individually consistent with r.e. [quasi-] conservativeness, their conjunction is not. Part of the explanation is this: without positive introspection, an r.e. but non-recursive theory R can count as satisfying negative introspection by falsely equating the agent's theory with a recursive subtheory of R; the idea behind the clause for $\theta\Box\alpha$ in the proof of Proposition 34.17 is to use PC as such a subtheory. That R satisfies negative introspection by making false equations is crucial, for KE[T] is S5 itself. Although both negative introspection and truthfulness are individually consistent with r.e. [quasi-] conservativeness, their conjunction is not.

Related Non-computational Constraints

Although r.e. conservativeness and r.e. quasi-conservativeness are defined in computational terms, something remains when the computational element is eliminated. For given any [consistent] theory R in L, r.e. or not, we might require an epistemic logic to leave open the possibility that R is exactly the restriction of the agent's theory to L. On this view, an epistemic logic should impose no constraint beyond consistency on the agent's non-epistemic theorizing. Thus we define a modal logic Σ to be *conservative* if and only if for every theory R in L, $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$ for some maximal Σ-consistent set X. Σ is *quasi-conservative* if and only if for every consistent theory R in L, $\mathrm{L} \cap \Box^{-1}\mathrm{X} = \mathrm{R}$ for some maximal Σ-consistent set X. Equivalently, Σ is [quasi-] conservative if and only if for every [consistent] theory R in L, $\{\Box\alpha : \alpha \in \mathrm{R}\} \cup \{\neg\Box\alpha : \alpha \in \mathrm{L} - \mathrm{R}\}$ is Σ-consistent. We can assess how far r.e. conservativeness and r.e. quasi-conservativeness are specifically computational constraints by comparing them with conservativeness and quasi-conservativeness respectively.

Theorem 34.19 A prenormal modal logic Σ is quasi-conservative if and only if for no $n \vdash_{\Sigma} \mathrm{Alt}_n$.

Proof ($\Rightarrow$) Suppose that $\vdash_{\Sigma} \mathrm{Alt}_n$. Put $X = \{\Box\alpha : \alpha \in \mathrm{PC}\} \cup \{\neg\Box\alpha : \alpha \in \mathrm{L} - \mathrm{PC}\}$. For all $i \leq n$, not $\vdash_{\mathrm{PC}} \wedge\{p_j : j < i\} \supset p_i$, so $\neg\Box(\wedge\{p_j : j < i\} \supset p_i) \in \mathrm{X}$. Hence $\mathrm{X} \vdash_{\Sigma} \neg\mathrm{Alt}_n$, so $\mathrm{X} \vdash_{\Sigma}$ is $\bot$. Since PC is a theory in L, Σ is not quasi-conservative. ($\Leftarrow$) Suppose that R is a consistent theory in L and $\{\Box\alpha : \alpha \in \mathrm{R}\} \cup \{\neg\Box\alpha : \alpha \in \mathrm{L} - \mathrm{R}\}$ is not $\Sigma -$ consistent. Thus for some $\alpha_0, \ldots, \alpha_m \in \mathrm{R}$ and $\beta_0, \ldots, \beta_n \in \mathrm{L} - \mathrm{R}$ (such β_i exist because R is consistent), $\vdash_{\Sigma} \wedge\{\Box\alpha_i : i \leq m\} \supset \vee\{\Box\beta_i : i \leq n\}$. Let $i \leq n$; since $\alpha_0, \ldots, \alpha_m \in \mathrm{R}$, $\beta_i \in \mathrm{L} - \mathrm{R}$ and R is a theory, it follows that for some valuation υ_i of L onto $\{0, 1\}$ (where $\upsilon_i(\bot) = 0$ and $\upsilon_i(\gamma_1 \supset \gamma_2) = 1$ just in case $\upsilon_i(\gamma_1) \leq \upsilon_i(\gamma_2)$), $\upsilon_i(\alpha_j) = 1$ for all $j \leq m$ and $\upsilon_i\ (\beta_i) = 0$. Put $\upsilon_{n+1} = \upsilon_0$. Set $\delta_i = \wedge\{p_j : j < i\} \wedge \neg p_i$ for $i \leq n$ and $\delta_{n+1} = \wedge\{p_j : j \leq n\}$. Let σ be the substitution such that for all j, $\sigma\ p_j = \vee\{\delta_i : \upsilon_i(p_j) = 1,\ i \leq n + 1\}$. Since Σ is closed under US, $\vdash_{\Sigma} \wedge\{\Box\sigma\alpha_i : i \leq m\} \supset \vee\{\Box\sigma\ \beta_i : i \leq n\}$. We can prove by induction on the complexity of γ that for all $\gamma \in \mathrm{L}$ and $i \leq n + 1$, if $\upsilon_i(\gamma) = 1$ then $\vdash_{\mathrm{PC}} \delta_i \supset \sigma\gamma$ and if $\upsilon_i(\gamma) = 0$ then $\vdash_{\mathrm{PC}} \delta_i \supset \neg\sigma\gamma$. Basis: Immediate by definition of σ, for $\vdash_{\mathrm{PC}} \delta_{\mathrm{i}} \supset \neg\delta_k$ whenever $i \neq k$. Induction step:

Routine. Now for $i \leq n+1$ and $j \leq m$, $\upsilon_i(\alpha_j) = 1$, so $\vdash_{\mathrm{PC}} \delta_i \supset \sigma\alpha_j$; since $\vdash_{\mathrm{PC}} \vee\{\delta_i : i \leq n+1\}$, $\vdash_{\mathrm{PC}} \sigma\alpha_j$, so $\vdash_{\mathrm{PC}} \top \supset \sigma\alpha_j$. Hence by prenormality $\vdash_\Sigma \Box\top \supset \Box\sigma\alpha_j$ and so $\vdash_\Sigma \Box\sigma\alpha_j$. Thus $\vdash_\Sigma \vee\{\Box\sigma\beta_i : i \leq n\}$. Moreover, for each $i \leq n$, $\upsilon_i(\beta_i) = 0$, so $\vdash_{\mathrm{PC}} \delta_i \supset \neg\sigma\beta_i$, so $\vdash_{\mathrm{PC}} \sigma\beta_i \supset (\wedge\{p_j : j < i\} \supset p_i)$, so $\vdash_\Sigma \Box\sigma\beta_i \supset \Box\ (\wedge\{p_j : j < i\} \supset p_i)$. Thus $\vdash_\Sigma \mathrm{Alt}_n$.

Proposition 34.20 A prenormal modal logic Σ is conservative if and only if Σ is quasi-conservative and not $\vdash_\Sigma \Diamond\top$.

Proof Like Proposition 34.1 with ‘r.e.’ omitted.

Thus S5 is a quasi-conservative normal modal logic which is not r.e. quasi-conservative; K4E is a conservative normal modal logic which is not r.e. conservative. Most of the examples given above of logics which are not r.e. [quasi-] conservative are [quasi-] conservative. It is the distinctively computational requirements of r.e. quasi-conservativeness and r.e. conservativeness which those logics fail to meet.

Corollary 34.21 Every r.e. quasi-conservative prenormal modal logic is quasi-conservative; every r.e. conservative prenormal modal logic is conservative.

Proof From Proposition 34.1, Corollary 34.6, Theorem 34.19 and Proposition 34.20.

Although quasi-conservativeness exceeds r.e. quasi-conservativeness in requiring an epistemic logic to leave open the possibility that the restriction of the subject’s theory to the language L is any given non-r.e. theory in L, this requirement is met by any epistemic logic which leaves open the corresponding possibility for every consistent r.e. theory in L.

Conclusion

Our investigation has uncovered part of a complex picture. The line between those modal logics weak enough to be r.e. conservative or r.e. quasi-conservative and those that are too strong appears not to coincide with any more familiar distinction between classes of modal logics, although a solution to the problem left open in the section “Some non-r.e. quasi-conservative logics” about the converse of Theorem 34.3 might bring clarification. What we have seen is that some decidable modal logics in general use as logics of knowledge (such as S5) or belief (such as KD45 and K45) when applied in generalized settings impose constraints on epistemic agents that require them to exceed every Turing machine in computational power. For many interpretations of epistemic logic, such a constraint is unacceptably strong.

The problem is not the same as the issue of logical omniscience, since many epistemic logics (such as S4 and various provability logics) do not impose the unacceptably strong constraints, although they do impose logical omniscience. Interpretations that finesse logical omniscience by building it into the definition of the propositional attitude that interprets the symbol $\Box$ do not thereby finesse the computational issue that we have been investigating. Nevertheless, the two questions are related, because the deductive closure of a recursively axiomatised theory is what makes its theorems computationally hard to survey. In particular, it can be computationally hard to check for *non*-theoremhood, which is what negative introspection and similar axioms require. In fact, negative introspection by itself turned out not to impose unacceptable computational requirements (Corollary 34.18), but its combination with independently more plausible axioms does so. Perhaps the issues raised in this paper will provide a more fruitful context in which to discuss some of the questions raised by the debate on logical omniscience and bounded rationality.

The results proved in the paper also suggest that more consideration should be given to the epistemic use of weaker modal logics that are r.e. conservative or quasi-conservative. The plausibility of correspondingly weaker axioms must be evaluated under suitable epistemic interpretations. Weaker epistemic logics present a more complex picture of the knowing subject, but also a more nuanced one, because they make distinctions that stronger logics erase. We have seen that the more nuanced picture is needed to express the limits in general cognition of creatures whose powers do not exceed those of every Turing machine.

Acknowledgements Material based on this paper was presented to colloquia of the British Society for the Philosophy of Science and the Computer Science Laboratory at Oxford. I thank participants in both for useful comments.

References

Blok, W. J. (1980). The lattice of modal logics: An algebraic investigation. *Journal of Symbolic Logic, 45*, 221–236.

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

Craig, W. (1953). On axiomatizability within a system. *Journal of Symbolic Logic, 18*, 30–32.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.

Makinson, D. C. (1971). Some embedding theorems in modal logic. *Notre Dame Journal of Formal Logic, 12*, 252–254.

Shin, H. S., & Williamson, T. (1994). Representing the knowledge of Turing machines. *Theory and Decision, 37*, 125–146.

Skyrms, B. (1978). An immaculate conception of modality. *The Journal of Philosophy, 75*, 368–387.

Williamson, T. (1996). Self-knowledge and embedded operators. *Analysis, 56*, 202–209.

Williamson, T. (1998). Iterated attitudes. In T. Smiley (Ed.), *Philosophical logic* (pp. 85–133). Oxford: Oxford University Press for the British Academy.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.