Relevant Alternatives in Epistemology and Logic

Peter Hawke

Abstract The goal of the current paper is to provide an introduction to and survey of the diverse landscape of *relevant alternatives theories of knowledge*. Emphasis is placed throughout both on the abstractness of the relevant alternatives approach and its amenability to formalization through logical techniques. We present some of the important motivations for adopting the relevant alternatives approach; briefly explore the connections and contrasts between the relevant alternatives approach and related developments in logic, epistemology and philosophy of science; provide a schema for classifying and studying relevant alternatives theories at different levels of abstraction; and present a sample of relevant alternatives theories (contrasting what we call question-first and topic-first theories) that tie our discussion to ongoing debates in the philosophical literature, as well as showcasing techniques for formalizing some of the important positions in these debates.

Key words: Relevant alternatives theory; epistemic relevance; epistemic closure; epistemic logic; questions; subject matter.

1 Introduction

The aim of the current paper is to introduce the reader to the relevant alternatives (RA) approach to the theory of knowledge and provide some indication of the complex landscape such theories inhabit.

One important theme that we emphasize throughout is the breadth and versatility of the RA approach - at least in the very general form we expound and develop it here. Indeed, the diversity of the existing theories of knowledge that fall under the

Peter Hawke
Stanford University, Department of Philosophy,
Building 90, 450 Serra mall, Stanford University, Stanford, CA 94305-2155,
e-mail: phawke@stanford.edu

RA banner - many of which we will meet in this paper, notably in sections 2.8 and 4 - bears testimony to the *abstractness* of our basic RA framework.

Another important theme is that RA theory, in its many guises, is typically amenable to study using precise formal methods. The RA approach is, therefore, not only a unifying framework for diverse, nuanced and intriguing philosophical theories of knowledge (encompassing, as should become apparent, a significant bulk of important recent developments for this ancient philosophical inquiry), but is also a notable site for the interaction between epistemology and logic. This interaction extends fruitfully in both directions [22, 24, 28, 29, 30]: logical techniques allow the RA theorist to operate at an unusual level of technical precision when framing rival positions and their consequences, informing the philosophical discussion in a manner that goes beyond mere window-dressing; while RA theory is a source of novel, sophisticated variants of epistemic logic, worthy of detailed logical study in their own right.

It is *not* a goal of the current paper to defend or attack any particular RA theory, or even the RA approach as a whole. Rather: it is to awaken in the reader an interest in RA theory as a venue for both epistemology and logic, illustrate the scope and dimensions of the RA approach and broach interesting questions for the RA theorist.

In the next section, I introduce the reader to the spirit of RA theory and review the motivation for this general approach, citing, for instance, some compelling linguistic considerations and the idea that RA theory captures the 'common man' response to the problem of cartesian skepticism. In addition, we briefly draw out some connections and contrasts between the RA approach and similarly themed discussions in the logic, epistemology and scientific methodology literature. In the third section, I propose a series of basic 'choice points' for the RA theorist. The leading claim here is that any *particular*, 'concrete' theory of knowledge that counts as an RA theory is essentially the product of settling each choice point. Hence, our list of choice points, it is suggested, offers a basic schema for *classification* of RA theories and provides a tool for studying RA theory at different levels of abstraction (where a higher level of abstraction corresponds to leaving more choice points open). We discuss each choice point in turn and briefly mention techniques for formalizing some of the potential paths associated with each choice point that the RA theorist can follow. In the fourth section, I exhibit a number of RA theories, suitably formalized using logical semantics, and classify them according to the schema from section 3. I associate these theories with concrete proposals from leading contemporary writers in the philosophical literature - specifically, Jonathan Schaffer and Stephen Yablo - and connect our discussion with important recent philosophical debates. With that, we conclude.

2 RA Theory: Its Nature and Motivation

2.1 The Slogan

The spirit of RA theory is quickly captured by the following slogan:

> In order for *S* to know that *P*, *S* need only have evidence that rules out all of the *relevant* alternatives to *P* (that is, *S* need *not* have evidence that rules out *all* of the alternatives to *P*).

Intuitively, one can think of an alternative to *P* as a circumstance that conflicts with *P* (we deliberately here use 'circumstance' in a vague and intuitive way, neutral between 'proposition', 'claim', 'state of affairs', 'possible world' and so forth). An alternative to the circumstance that Bush won the election is that Gore won the election. An alternative to the circumstance that Alan Turing was born in 1913 is that he was born in 1914 (this alternative to the circumstance in question is in fact not the case), and another is that he was born in 1912 (this alternative is in fact the case)[1].

Also intuitively, one may approach the notion of *ruling out* by way of that of *evidence*: the function of evidence is to *rule out* alternatives. If I were a detective trying to solve a murder case and my hypothesis is that the perpetrator was the butler, evidence that establishes that the maid has an alibi is *significant* evidence as it rules out the alternative that the maid is the murderer. Of course, 'ruled out' has an intuitively *strong* reading (to be contrasted with, say, 'unlikely') that seems befitting of association with the term 'knowledge'.

To say then that knowledge of *P* involves acquiring evidence good enough to rule out all of the alternatives to *P* has, to many ears, the air of a platitude. The RA theorist thinks that this saying is only half-right, however: coming to know involves ruling out only *select* alternatives, those that are *(epistemically) relevant*[2].

Obviously, what 'relevance' comes to is a key concern when judging an RA theory. We say more about relevance later (specifically, section 3.4). It is worth immediately stoking some intuitions, however: irrelevant alternatives are ones that are (in context) "far-fetched"; are not to be "taken seriously" when making judgements concerning knowledge; are rightly "ignored", in some sense, when it comes to matters epistemic. A suggestive example: to know that my *left* neighbour's dog is barking, I need to rule out that the barking sounds I hear are not emanating from the direction of my *right* neighbour's house. But it might seem that I do *not* need to rule out the bizarre possibility that my left neighbour's dog has been kidnapped and the kidnappers left behind a recording device to play back the sound of a dog's bark as an elaborate ruse.

---

[1] For this paper, we set aside the tricky question as to in what sense circumstances that are necessarily the case or not the case (such as Goldbach's conjecture) can be thought to have alternatives - at least for the purposes of inquiry.

[2] Note that in section 2.9 we discuss a tradition in the epistemology literature that focuses on a notion of "epistemic relevance" that arises from initial concerns quite distinct from those of the RA theorist. The overlapping terminology is no doubt a potential source for confusion, though hopefully not in the current paper.

It should be emphasized from the outset how little content we initially commit to in our introduction of the notion of "relevance" (indeed, this trend continues as we discuss the basic motivations for the RA approach in the coming sections). *All* that we begin with is the idea that (and some reason to think that) some alternatives are relevant to the evaluation of an agent's knowledge, and others are not. In particular, we do *not* in the initial statement of the RA approach commit to the idea that relevance is a matter of rationality, irrationality or arationality; and we do not commit to the idea that relevance is a function of context, conversational or otherwise. Our statement of the RA approach leaves these questions open.

2.2 Motivating the RA Approach

Why be an RA theorist? In the following sections, we present a number of important motivations (many of which, the reader might note, are related, though still worth separating out). Let us begin with an overview.

- There is *striking linguistic data*, concerning our ordinary usage of epistemic claims, that seems to support RA theory.
- The RA approach provides a unique and compelling reply to *cartesian skepticism*.
- More generally, RA theory provides a universal strategy for dealing with *underdetermination problems*.
- RA theory is suggested by our intuitive reaction to *Goldman-Ginet barn cases*.
- RA theory has theoretical value (for contextualists and others) as an attractive and convenient tool for *measuring epistemic standards*.

2.3 Suggestive Linguistic Data

Two kinds of purported linguistic data have been used (in concert) to support the RA approach. First, linguistic data seems to indicate a *fallibilist* aspect to our ordinary knowledge concept. That is, it seems that ordinary agents will sometimes happily attribute knowledge to themselves (or others), but, *if pressed*, will concede that certain possibilities for error are compatible with the available evidence. Second, linguistic data seems to indicate an *infallibilist* aspect to our ordinary knowledge concept. That is, ordinary agents seem uncomfortable to state the conjunction of a knowledge claim with an explicit acknowledgement of live possibilities of error. These points are emphasized by both Dretske [15] and Lewis [36], following in the footsteps of Unger [50] (though it should be noted that [50], a thoroughgoing defense of skepticism, hardly supports an RA approach).

We may immediately note the tension between the seeming fallibilist and infallibilist tendencies of ordinary knowledge ascription. As we shall see, one alleged advantage of the RA approach is its seeming capacity to resolve this tension.

To illustrate the fallibilist tendency, we (ab)use an influential case. Fred Dretske [13] is famous for pointing out that, under ordinary circumstances (using ordinary visual evidence), one will seem perfectly happy to say that one knows that the animal one sees at the zoo - in the zebra enclosure - is a zebra. One will be *less* happy, it seems, to say that one knows that the animal is not a mule painted to appear like a zebra. The ordinary visual evidence does not seem to settle the latter issue. Also famously, Dretske uses this example as a counter-example to the claim that knowledge is closed under known entailment: one can know P, know that P entails Q, put "two and two together" and yet not know that Q.

The legitimacy and exact diagnosis of this purported linguistic data, and its consequences for the truth of the closure principle, are controversial [51, 38]. For our purposes, however, note that the description of the example may be slightly altered in a telling way, by weakening the proposed judgement concerning the "painted mule" possibility: under ordinary circumstances - it is plausibly suggested - one is happy to say one *knows* the enclosed animal is a zebra, yet, if pressed, will be hesitant to add that one has evidence that *rules out* that the animal is a painted mule. Yet being a painted mule is an alternative to being a zebra. Thus, it appears we have everyday linguistic data to the effect that we are often willing to ascribe knowledge of P, yet will quickly concede the limitations of the available evidence when it comes to ruling out *certain* alternatives to P[3].

As has been pointed out by critics of the RA approach [51, 52], this modest reading of our intuitions in the zebra case may support fallibilism in general, but does not support the RA approach in particular. For there are other prominent fallibilist approaches to the theory of knowledge. Consider, for instance, a form of Bayesian that holds that knowledge of P is essentially a matter of not-P being *sufficiently improbable* on the evidence. Such a Bayesian, it seems, is a rival to the RA theorist, seeing no need for evidential support (in its role as constraining the space of possibilities) to be supplemented with an independent notion of "relevance".

This form of Bayesianism, however, does not seem as effective in accounting for the *infallibilist* tendencies in our ordinary knoweldge ascriptions. Ordinary speakers, it seems, feel uncomfortable in making or accepting claims along the following lines: "I know that P, though not-P might well be the case"; "I know that P, yet my evidence does not vouchsafe certainty that P"; "I know that P, though not-P remains a live possibility". Lewis sums up the sentiment effectively:

> "If you claim that S knows that P, and yet you grant that S cannot eliminate a certain possibility that not-P, it certainly seems as if you have granted that S does not after all know that P. To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* contradictory" [36, pg.549, his emphasis].

The aforementioned Bayesian approach seems an awkward fit with infallibilism. According to this account, one can know P when the probability bestowed on P by the evidence meets an appropriate threshold. But if this threshold is less than 1,

[3] Throughout this section, the critical reader may well want to emphasize the use of the word *appearance* here in the absence of a proper empirical investigation of these purported linguistic facts.

then the Bayesian is committed to the possibility that an agent may know that P and yet not-P has non-zero probability and, so, is compatible with (if unlikely on) the evidence.

The RA theorist, on the other hand, has a trick to play. She can account for our fallibilist tendencies: if P is known and yet an alternative A to P is identified as uneliminated by the evidence, then A, the RA theorist proposes, is (or, at least, *was* in context) an irrelevant alternative to the evaluation of knowledge of P. On the other hand, the RA theorist can account for our infallibilist tendencies. She can essentially agree with the ordinary assessment that "to know is to leave *no* possibility for error". But what is left *implicit* in such a saying, the RA theorist proposes, is that the possibilities being quantified over are only the *relevant* ones in that conversational context [4].

2.4 The RA Strategy Against Skepticism

Consider the following argument for a skeptical conclusion:

P1. To be a handless brain-in-a-vat is an alternative to having hands.
P2. The evidence in my possession is not sufficient to rule out that I am a handless brain-in-a-vat.
P3. In order to know P, one needs to have evidence that rules out all alternatives to P.
C. **Therefore:** I do not know that I have hands.

This argument is valid, and P1 and P2 might strike one as undeniable (to deny them, it might be said, is simply not to correctly appreciate the nature of the brain-in-vat scenario). To resist skepticism, there seems only one way out: deny P3. Of course, this is simply to embrace the RA slogan.

I note that, at least in my experience, something along these lines is a common response from the layman (i.e. non-philosophers) when presented with the threat of skepticism. The reaction, it seems, is to deride the brain-in-vat scenario as far-fetched and otherwise *irrelevant* to our ordinary epistemic concerns. Such a reaction seems particularly apt when a practical application of everyday knowledge is afoot. It is in no way an adequate response to the question "do you know where I left my keys?" to say "no, for I cannot rule out that my senses are being deceived by an evil demon". To the extent that she is willing to take the layman as a competent user of the knowledge concept, the RA theorist finds this reaction telling[5].

---

[4] Our ordinary infallibilist tendencies have in fact been used as a weapon in *internal* debates among RA theorists, suggesting the possibility that some versions of RA theory are better suited to account for these tendencies than others. For instance, DeRose [12] influentially criticizes Dretske's version of RA theory as incorrectly predicting that so-called *abominable conjunctions* - notably "S knows that she has hands and S does not know that she is a handless brain-in-vat" - are felicitous in ordinary conversational contexts.

[5] Of course, this may be taken as further linguistic data, in the spirit of that from section 2.1.

2.5 RA Theory as a Response to Under-determination Problems

Let us generalize the previous point of motivation. Cartesian skepticism, at least in certain forms, is an instance of a larger class of problems that may be called *under-determination problems*. An under-determination problem has the following form: it is obvious that we know that P, yet, on close inspection, our supposed evidence for P seems just as compatible with some (maybe odd, but logically possible) alternative Q. Another prominent under-determination problem: Humean skepticism, where our sensory evidence of particulars seemingly under-determines the general knowledge we tend to hold upon its basis. In general, under-determination has proven a pressing issue in philosophy of science [48].

An RA theory embodies a universal strategy for dealing with under-determination problems: simply establish that any deviant alternatives are properly classified as irrelevant (whatever this classification comes to).

2.6 RA Theory by Way of the Goldman-Ginet Barn Case

Along with the Gettier examples and Dretske's painted mule example, the Goldman-Ginet barn case [19] has been a particularly influential example in the contemporary epistemology literature. Suppose subject S clearly observes what is in fact a (genuine) barn out of her car window, as she drives by. Does she know that it is a barn? Our reaction to this question will depend, the example seems to show, on whether S is driving through a county in which the only objects that look like barns to the casual observer are, in fact, barns (in which case, she does know), or if she is in the unusual situation where there are as many barn facades ("fake barns") around as real barns (in which case, she does not).

What exactly does the barn case teach us? The RA theorist may point out the following: it seems to demonstrate that it is possible that S has *exactly the same evidence* in states s and s' (not to mention the same beliefs), and yet S knows that P in s and does not know that P in s' (where P is true in both s and s'). The difference, the RA theorist will urge us: different alternatives to P are relevant in one case than in the other, and, in particular, S does not have sufficient evidence to rule out an alternative (that the object is a barn facade) that happens to be irrelevant in s and happens to be relevant in s'[6].

[6] Note that the barn case can be seen to teach a similar lesson to consideration of cartesian skepticism: that one can know something even though one has not ruled out *all* alternatives. However, the barn case potentially teaches us something more: that what counts as a relevant alternative can *vary* with the circumstances: the possibility of fake barns may be properly ignored, by knowledge ascribers, under one set of circumstances, but is not properly ignored in another.

2.7 RA Theory and Epistemic Standards

A great number of authors in the recent epistemology literature have defended some version of the idea that the *epistemic standards* that an agent needs to meet in order to know that P can vary from context to context. What is chiefly debated, amongst such authors, is *which* context determines the relevant standards: is it that of the *subject* to whom knowledge is potentially attributed [49], that of the *speaker* who is performing the attribution [10, 12, 36], or that of an *assessor* potentially different to both speaker and subject [39]?

Whichever view one takes, such perspectives on the semantics of knowledge claims are a natural fit with RA theory. For how are we to understand the idea of an *epistemic standard*? A natural suggestion is that a variation of epistemic standards consists in a variation of the amount of alternatives that need to be ruled out: a higher standard involves a larger amount of relevant alternatives.

Thus, it is a short path from accepting that epistemic standards vary by context to an acceptance of some form of RA theory. Arguments for the former may therefore, with the right massaging, be taken as support for the latter[7].

2.8 The History of RA Theory

We have thus far discussed RA theory and its motivation as if it exists in a vacuum. In fact, the list of active and explicit defenders of RA theory in the literature is long and varied: Dretske 1970 [13], Goldman 1976 [19], Dretske 1981 [15], Luper 1984 [37], Lewis 1996 [36], Cohen 1988 [10], Heller 1989 [25], Pritchard 2012 [42], Lawlor 2012 [34], Holliday 2015 [30]. As I will explain in a moment, it is reasonable to add to this list Austin 1946 [2] and Nozick 1981 [41]. Let us briefly delve into some of the history of support for RA theory.

J.L. Austin is notable for making especially early remarks in the direction of an RA theory. His suggestions are discussed and expanded at length by Lawlor [34].

Fred Dretske [13, 15], however, may be singled out as fully initiating the ongoing discussion of RA theory. Dretske's view, somewhat obliquely presented in his initial paper, is roughly as follows: for agent S to know that (true) P, S must believe P on the basis of a conclusive reason R, where R being conclusive means that: if P had not held, then neither would R have held. Averting to the standard ideas in the literature on the semantics for counter-factual conditionals we can say: knowledge requires that in the nearest worlds to actuality in which P is false, so too is R false. We may say then that for Dretske, roughly, an alternative Q to P is relevant just in case it holds at the nearest worlds in which P is false, and Q is ruled out just in case those worlds are incompatible with the agent's reasons (evidence/information).

[7] For a more careful defence of the 'alternatives' approach to capturing the relevant parameter that shifts across contexts, see [45].

On Dretske's view then, the relevance of Q is relative to the proposition being considered as an object of knowledge: Q might well be relevant relative to one proposition and irrelevant relative to another.

We might contrast this theory to that of Lewis [36], another prominent and influential RA theorist: roughly, according to Lewis, relevance is determined by a complex set of rules operating on the conversational context in which knowledge attributions may be made. Thus, relevance is relative to a conversational context common to all propositions, as opposed to a specific proposition evaluated for knowledge: if the context is held fixed, proposition Q is fixed as relevant (or irrelevant, as the case may be), no matter which proposition it is contrasted to as an alternative.

I re-emphasize an important point for the current paper brought out by these observations: on occasion in the literature, the label "relevant alternatives theory" is very closely associated with Dretske's theory in particular (and Lewis, for instance, is classified, in contrast, as a "contextualist"). As should by now be evident, in this paper we use the term "relevant alternatives theory" in a liberal and broad manner that encompasses a wide range of views. Indeed, given its structural similarities to Dretske's view (simply replace talk of "having a conclusive reason" with "having a sensitive belief") we could happily class Nozick's well-known tracking theory of knowledge (and its variations) under the RA banner. As we will see in section 4, we can also convincingly fit recent work by Schaffer and Yablo under the RA banner, though again these authors do not tend to self-describe their views with this label. In our view, the unifying generality and abstractness of the RA approach is part of its appeal as an object of study.

2.9 Connections and Contrasts: Relevance Logic, "Epistemic Relevance" and Scientific Methodology

We close this section with some brief discussion of the potential connections and contrasts between the RA approach and other salient developments in the epistemology and logic literature. Our aim is to achieve some sense of the theoretical promise of the RA approach (insofar as it can intergrated and unified with similarly motivated concerns in other strands of the literature) while also being sure to *distinguish* the concerns of the RA theorist from sometimes only superficially similar issues.

(For readers keen to immediately dig into more nitty-gritty features of the RA approach, note that this section may be skipped without any significant break in the flow of the paper).

2.9.1 Relevance Logic

Begin with the well-developed field of *relevance logic* [1, 6, 40]. In brief, what animates this area of logic is a desire to build (technically and philosophically sound)

logics that avoid endorsing so-called "fallacies of relevance" as valid. In particular, the relevance logician is concerned to avoid two counter-intuitive results of classical logic: that any sentence is a valid consequence of contradictory premises, and that a necessary truth is a valid consequence of any set of premises whatsoever. The difficulty with these results, the relevance logician claims, becomes immediately evident when we consider cases where the premises and conclusion are *irrelevant to each other* insofar as they concern disjoint *subject matter*: it does not follow from the claim that the moon is both made of green cheese and not that Barack Obama is president of the USA (nor, for that matter, does it follow from 2+2=5). Further, it does not follow from the fact that Berlin is the capital of Germany that either it is raining in London or it is not (nor, for that matter, that 2+2=4).

In sum then (though we place ourselves at risk of over-simplifying), the relevance logician has two concerns: (i) to offer an account of when one proposition is "relevant" to another (which, for all we have said, appears to amount to accounting for what it means to say that two propositions overlap in subject matter) and (ii) an integration of this account into a logical system, to the effect that only *relevant* conclusions are valid consequences of a set of premises. The concerns of the relevance logician and a RA theorist overlap, therefore, to the extent that (i) and (ii) are pertinent to the RA theorist in question.

Is (i) pertinent to an RA theorist? This will depend on whether the RA theorist and relevance logician mean the same thing by "relevance". Since they are motivated by different starting points (which alternatives can be "properly ignored" when evaluating knowledge claims versus what intuitively follows from what) there is no guarantee that there will in general be a convergence here. Indeed, there is a quick argument that "relevance" as deployed by a standard RA theorist must have a different sense (or at least application) than that deployed by the relevance logician. For: suppose we follow the standard line (we return to this in section 3.5) and say that proposition A is an alternative to P just in case P entails $\neg A$. Now, the RA theorist wishes to draw a *distinction* between relevant and irrelevant alternatives to P. But this distinction seems to rely on the claim that *both* kinds of alternative are logically related to P. *Both* are "relevant", therefore, in the sense of the relevance logician.

Nevertheless, we discuss in section 4 *topic-first RA theorists* that attempt to account for "relevance" in terms of subject matter. For such RA theorists, agreement on considerations of relevance might be sturdy enough for a useful dialogue with the results of relevance logic.

Is (ii) pertinent to an RA theorist? On the face of it, the answer is 'yes'. Suppose our RA theorist has settled on an account of relevance. There is then clear theoretical interest for her in developing a logical system where relevance is preserved across the proposed logical consequence relation. What is not so clear, however, is that the results of relevance logic provide a general enough framework for carrying out this job for arbitrary RA theorists, since, again, relevance logic is typically associated with a notion of relevance closely tied to preservation of *subject matter*. To the extent that interest in the RA approach fuels an interest in a diversity of accounts of

relevance (see section 3.4), it motivates a wider scope for relevance logic than mere attention to the interaction of consequence and subject matter.

In total: the basic concerns of relevance logic, at the very least, indicate an intriguing notion of relevance tied intimately to that of subject matter, an obvious matter of interest to the RA theorist. Beyond this, dialogue between relevance logic and the RA approach presents an intriguing, possibly fruitful but certainly subtle affair. Of course, our remarks are tentative: the relationship between relevance logic and RA theory deserves a more careful discussion.

2.9.2 Epistemic Relevance Between Evidence and Hypothesis

Let us now turn to a second tradition in philosophy in which the term "relevance" has received prominence. Here, the focus has been on when a piece of evidence is *relevant* to the evaluation of a hypothesis. There is therefore a parallel with the concerns of the relevant logician: while the relevant logician is concerned with when a conclusion genuinely follows from its premises, so one who investigates "epistemic relevance" in the present tradition is concerned with when evidence is genuinely a *reason* to accept (or reject) a hypothesis. The traditional starting point in this investigation has been a probabilistic account that states the following: evidence E is relevant to hypothesis H just in case the conditional probability of H given E is different to the (prior) probability of H. Discussion in the literature - initiated chiefly by Keynes and Carnap - has essentially developed as a series of refinements of this basic idea [31, 7, 18].

Analogously to the case of relevance logic, the discussion of "epistemic relevance" hinges on two basic concerns: (i*) what is the correct account of the relevance at issue? (ii*) How is this account to be integrated into a theory of evidential support? Once again, the extent to which the discussion of this sense of epistemic relevance relates to the concerns of the RA theorist depends on the extent to which answers to i* and ii* bear on these concerns.

On a first pass, we may make observations similar to those made with respect to the relationship between the RA approach and relevance logic. With respect to i*: the notion of relevance at work in the discussion of "epistemic relevance" is of interest to the RA theorist insofar as it represents, surely, *one* intriguing candidate for the notion of relevance the RA theorist believes can be identified as at work in the theory of knowledge (namely, a candidate that appeals to notions of *probability* and *independence* as crucial features). It remains to be seen, however, how far such a version of RA theory could be developed with plausibility. With respect to ii*: again, the RA theorist certainly ought to have interest in any general techniques for integrating an account of relevance into a theory of reasoning or evidential support (perhaps in aid of a *relevant alternatives theory of justification* that underlies the RA theory of knowledge). The apparent focus in the "epistemic relevance" literature on a quite specific notion of relevance does not inspire hope, however, that very *general* tools for such integration are to be found there. Once again, however,

our remarks cannot be understood as anything other than preliminary, and clearly a deeper investigation is a worthwhile task.

2.9.3 Methodology of Science

Various strands in the literature on the epistemology and methodology of science have, it seems to me, a notable *prima facie* affinity to the ideas animating the RA theorist (not forgetting, of course, the motivation for the RA approach as a universal solution to under-determination problems, discussed in section 2.5). We sketch a few such points.

Begin with the plausible idea that the *goal* of scientific inquiry is knowledge. If then we agree with the RA theorist that knowledge is always relative to a set of relevant alternatives, it seems that we should conclude that the methods of scientific inquiry - geared towards producing such knowledge - themselves operate against the backdrop of a set of relevant alternatives. If so, we should expect a notion of "relevant alternative" to play a role in both the *context of justification* and *context of discovery* of scientific hypotheses.

Indeed, ideas of roughly this ilk have received considerable attention in the literature. A linchpin of discussion in the philosophy of science literature of the last few decades has been Kuhn's proposal that major developments in the history of science amount to revolutionary upheavals brought about by a fundamental shift in underlying "paradigm" for normal science [33]. We need not here become engrossed in the substantive or scholarly issues connected to Kuhn's work. We need only note a potential for an RA approach to offer tools to understand and investigate such revolutionary shifts: for an RA theorist, a change in paradigm, it might be suggested, involves a major shift in the space of the *relevant hypotheses* that a normal scientist must seek to select between.

More specifically, let us consider the *context of justification*. In recent years, there has been a revival of the idea that scientific justification essentially amounts to a process of *eliminative induction* (cf. [16, Ch.7]): roughly, given a space of relevant hypotheses H_1 through H_n, a particular hypothesis H_i is supported by scientific inquiry just in case the available evidence rules out every competing hypothesis. One reason to initially find such an account of scientific methodology to be philosophically naive is to point out both the unwieldiness of the space of logically possible hypotheses, and the inability of our actual evidence to rule out any significant portion of this space (a version, of course, of *under-determination problem*). A successful RA theory, however, will presumably provide a notion of relevance that essentially defuses both problems. RA theory seems, therefore, a potential ally to the eliminative inductivist.

Turn now to the *context of discovery*. For the RA theorist, a natural way to understand the *discovery* of a new hypothesis is for that hypothesis - through whatever mechanism - to become *relevant* in the context of scientific inquiry. Whether this mechanism is rational or not will depend on the exact account of what relevance is, and how the space of relevant alternatives might change. In this connection, con-

sider Hintikka's recent work in developing a logic of discovery that, roughly, posits scientific discovery as essentially amounting to posing a question to a source of information in nature [27]. On such an approach, scientific inquiry depends crucially on the *questions* that are (implicitly or explicitly) asked by scientists. To the extent that *background questions* may therefore be understood as a source of *relevant alternatives* (see our discussion of *question-first RA theory* in section 4 of this article), we see yet another venue for potential convergence between RA theory and important debates in the methodology literature.

3 Choice Points for the RA Theorist

In this section, we make a start at developing the RA approach with technical precision.

First, we present a 'minimal' RA theory. This theory is minimal insofar as it operates at a high level of abstraction yet, we claim, captures the core elements of the approach. In the process, we lay out a logical language that we work with throughout the rest of the paper, and a basic semantics for this language.

The minimal approach is too abstract to engage fully with philosophical debate. Likewise, its abstractness precludes it from encompassing the interesting formal features of more concrete RA theories. To this end, we discuss the potential for considering precise RA theories with more content. With this in mind, we next list a number 'choice points' for the RA theorist, by which one may divide the family of RA theories into a large number of species. To settle each choice point is to arrive at a 'concrete' RA theory.

It is worth emphasizing two points concerning the philosophical motivation for the formal work that follows. First, as is often the case with highly abstract frameworks, minimal RA theory holds limited theoretical interest in itself. Rather, it is a unifying skeleton upon which to hang the features of more concrete RA theories. Nevertheless, an important philosophical point is attached to our presentation of a minimal theory: at its most abstract, the RA approach is *very general*, a point for critics to keep in mind when aiming for blanket objections to the approach. Our second point is an acknowledgement that, from a philosophical point of view, the utility of a move to the formalities of a logical approach - with its accompanying idealization and austerity - is to be judged by its pay-off for the perspicuous study and presentation of philosophically relevant results. We hope to demonstrate modest but genuine results along these lines in section 4.

3.1 An Epistemic Language

Let **At** be a set of atomic proposition letters. We work with the following logical language $\mathscr{L}$:

$$\varphi \ ::= \ p \,|\, \neg\varphi \,|\, \varphi \wedge \varphi \,|\, K\varphi \,|\, R\varphi \,|\, I\varphi \,|\, [\varphi]\varphi$$

where $p \in \mathbf{At}$.

The rest of the connectives are defined as usual. $K\varphi$ is intended to mean "the agent knows that φ". $R\varphi$ is intended to mean "φ is relevant". $I\varphi$ is intended to mean "the agent has the information that φ". The intended interpretation of $[\varphi]\psi$ is "after the set of relevant propositions is updated so as to be relative to φ, ψ is true". This last expression represents a *dynamification* of our logic [4] (and we fully intend to draw on the techniques of this area - which includes the likes of well-studied dynamic epistemic logics such as *public announcement logic* - in our development).

We may then define a *two-place* relevance operator: $\mathtt{R}(\varphi,\psi) ::= [\varphi]R\psi$. The aim here is to capture the idea that ψ is relevant (perhaps only) relative to proposition φ.

3.2 Minimal RA Theory

In what follows, $\mathscr{P}(A)$ refers to the power-set of set A.

Definition 1 (Minimal RA model). A *minimal RA model* is a tuple

$$\langle W, \{\mathbf{R}_w\}_{w\in W}, \{E_w\}_{w\in W}, \{*_w\}_{w\in W}, V\rangle$$

where,

- W is a set of *points of evaluation*. The reader may think of these as "possible worlds", subsets of which are "unstructured propositions".
- $\mathbf{R}_w \in \mathscr{P}(\mathscr{P}(W))$ is a set of sets of worlds i.e. a set of propositions. This is the set of *relevant propositions* at world w.
- $E_w \in \mathscr{P}(W)$ is a set of worlds i.e. a proposition. This is the agent's *total evidence* or *total information* at world w.
- $*_w$ is an *update operation* accepting a sentence $\varphi \in \mathscr{L}$ and returning an updated model we denote by $\mathscr{M} *_w \varphi$. We stipulate that the only distinction between $\mathscr{M}$ and $\mathscr{M} *_w \varphi$ lies in the relevant propositions.
- V is a valuation assigning atoms to worlds.

Given minimal RA model $\mathscr{M}$ and world w, define the set $\mathbf{U}_w$ as follows:

$$\mathbf{U}_w = \{A \subseteq W \,|\, A \in \mathbf{R}_w \text{ and } A \cap E_w \neq \emptyset\}$$

Call $\mathbf{U}_w$ the set of *uneliminated propositions* at w: the set of propositions that are both relevant and compatible with the agent's evidence at w.

Two remarks are in order. Though we use the 'worlds' terminology to talk about our points of evaluation, there is no technical necessity attached to this interpretation. One may equally well talk about scenarios, centered worlds, or so forth. (Though, it should be remarked, the totality of the propositional valuations associated with each world would make one hesitate to think of them as mere 'situations'.)

Second, we deliberately leave it vague how exactly to interpret E_w. Is this the evidence that the agent has access to *in principle*, though she may not in fact *have* all this evidence in her possession? Is this a conjunction of the individual pieces of evidence at the agent's disposal? If this is the agent's *information*, are we to understand E_w as being a *true* proposition i.e. $w \in E_w$? Indeed, settling these questions might indicate the unsuitability of representing E_w as a proposition, as opposed to, say, a set of (possibly incompatible) propositions. We say little to settle such questions in the present paper.

We now turn to semantics. As is typical in epistemic logic, it is worth bearing in mind that this semantics is best understood as describing an *idealized* agent. Idealized in what sense? For our purposes, we may understand our agent as follows: for our agent, relative to her information, there is no distinction between actual (explicit) and potential (implicit) knowledge. An ordinary human has much implicit knowledge relative to her information: knowledge that *could* be acquired by correct reasoning from the information and knowledge she already holds and yet, for whatever reason, she has not in fact acquired this knowledge. Our idealized agent has no such limitation.

Idealization raises the question: what is the relationship between our ideal agents and ordinary human beings? In particular, why think that a logical analysis of the one will shed light on the other? This is a subtle issue that deserves more discussion than we can give it here. One or two quick suggestions as to the relevance of ideal agents might prove useful to the reader, however. First, presumably, the study of any epistemic *limitation* of our ideal agents will have bearing on ordinary agents, since the former will also face that limitation. This observation seems particularly pertinent when it comes to under-determination problems, since, presumably, our ideal agents have no special advantage in terms of the empirical evidence at their disposal. Second, the move to idealization allows for elegant simplifications of certain issues. For instance, we will later meet certain proposed principles of epistemic closure that are difficult to state in full generality for ordinary agents, but simple to state in the case of ideal agents.

In what follows, read $[[\varphi]]_{\mathscr{M}}$ as:

$$[[\varphi]]_{\mathscr{M}} = \{w \in W \mid \mathscr{M}, w \vDash \varphi\}$$

Definition 2 (Minimal RA semantics). Given a minimal RA model $\mathscr{M}$, we define satisfaction at world w as follows:

- $\mathscr{M}, w \vDash p$ just in case $p \in V(w)$.
- $\mathscr{M}, w \vDash \neg\varphi$ just in case $\mathscr{M}, w \nvDash \varphi$.
- $\mathscr{M}, w \vDash (\varphi \wedge \psi)$ just in case $\mathscr{M}, w \vDash \varphi$ and $\mathscr{M}, w \vDash \psi$.
- $\mathscr{M}, w \vDash R\varphi$ just in case $[[\varphi]]_{\mathscr{M}} \in \mathbf{R}_w$.
- $\mathscr{M}, w \vDash I\varphi$ just in case $E_w \subseteq [[\varphi]]_{\mathscr{M}}$.
- $\mathscr{M}, w, \vDash K\varphi$ just in case $\{A \in \mathbf{U}_w \mid A \subseteq [[\neg\varphi]]_{\mathscr{M}}\} = \emptyset$.
- $\mathscr{M}, w \vDash [\varphi]\psi$ just in case $\mathscr{M} *_w \varphi, w \vDash \psi$.

Effectively, the clause for $I\varphi$ says that the agent has the information that φ just in case the agent's information entails φ. The clause for $K\varphi$ says the following:

the agent knows φ just in case there is no proposition that entails $\neg\varphi$ that is uneliminated i.e. both relevant and compatible with the evidence. The clause for $[\varphi]\psi$ simply says that such an expression is satisfied when ψ holds when the relevancy sets have been updated according to the $*_w$ operation with φ as input.[8].

A technical remark is in order. Our approach to the semantics of minimal RA theory clearly falls within the tradition of *neighbourhood semantics* for modal logic [8], where the truth clause for $\Box\varphi$ ("it is necessary that φ") amounts to: $\Box\varphi$ holds at world w just in case the set of worlds where φ holds is one of a set of "necessary propositions" associated with w. A technical elaboration of our proposed logic will therefore make use of the tools of neighbourhood semantics.

By now, the reader may well feel that we have left out something important in our RA account of knowledge. Presumably, knowledge is factive: if P is known, then P is true. What is more, knowledge implies belief - perhaps even *justified* belief. Yet none of these (some would say obvious) features are represented in our account of knowledge.

Incorporating these features into an RA account is a more subtle business than might first meet the eye [23]. We illustrate the issues by remarking on the factivity of knowledge.

One option for the RA theorist is simply to add an additional component to the truth condition for knowledge: in addition, it must be the case that φ is *true* at w. This is of course a structural feature of countless proposed theories of knowledge.

Such a maneuver might strike those that wish to understand attaining a knowledge state as offering a *guarantee* of truth as somewhat *ad hoc* and unsatisfying. More satisfying, it might seem, would be an account of knowledge such that the conditions on knowledge attainment *non-trivially entail* truth. Indeed, the independence of a 'truth condition' from the other conditions in a theory of knowledge might be exactly what allows for the construction of the familiar Gettier cases that have dogged epistemology [54].

It might therefore be seen as an advantage of RA theory that it provides tools for ensuring that truth is entailed by knowledge without the stipulation of an independent truth condition. For all that would be required is that: any true proposition at w is relevant at w; and the proposition E_w lives up to the title of 'information' by in fact being a *true* proposition. Both proposals have some appeal: it might seem strange to deem the truth as irrelevant, and it might seem natural to insist that E_w constitutes the agent's *basic evidence* (say, her memories and immediate sensory experience [36]) and that such evidence must be compatible with the actual world. Of course, these quick remarks do not settle the matter.

The foci of our discussion in the rest of this paper means that we can generally safely put aside the issue of truth and justified belief, so we will for the moment simply fail to propose a way of incorporating these features. The reader is right to recognize the gap, however.

---

[8] The reader will note that we make no mention of a notion of 'context' anywhere in this semantics. We gloss over the role of context, as follows: context may be thought of as settling the valuation V and, potentially, the set of relevant alternatives $\mathbf{R}_w$. Thus, context may be thought of as settling the model in question. We do not explore this thought in any detail here.

3.3 Choice Points

We now turn to a series of choice points the RA theorist must settle in order to fill out the minimal approach. We summarize the points here, before elaborating in the coming subsections (though we first pause to better note my intentions in enumerating this list).

An RA theorist ought to ultimately answer the following questions:

- What is relevance?
- What sort of thing is an 'alternative'?
- What is it to 'rule out' an alternative?
- What are the *primitive* objects of relevance from which we derive relevance of an alternative?
- Does the (ir)relevance of a claim only make sense in *contrast* to another claim?
- Interaction principles: is relevance a necessary condition on knowing? Is irrelevance a sufficient condition for knowing the denial?

Some notes about this list of choice points. First, I do not intend to be understood as claiming that these choice points are entirely *independent* of one another. Indeed, we will see some instances (in section 4) of how settling certain choices in a particular way constrains how other choices can be settled. Second, I do not claim that the manner in which an RA theorist can settle these choice points is entirely arational or arbitrary: there may well be good reasons for favouring one choice over another. Third, though I suspect this list is complete (in the sense that settling these issues produces a concrete RA theory), I will refrain from defending this point here.

3.4 Relevance

Perhaps the key philosophical matter that the RA theorist needs to settle is the question as to what *relevance* comes to.

This is no simple matter: the literature on RA theory displays a bewildering diversity of suggestive comments, but is light on detailed theories of relevance. Cohen [11, p.61] suggests that relevance is a matter of the psychology of the agents in conversation, "determined by some complicated function of speaker intentions, listener expectations, presuppositions of the conversation, salience relations etc". Heller [26] suggests that relevance is a matter of similarity to the actual world, where the similarity relation is itself settled partially by psychological facts - intentions, salience and so forth - of the speakers in context. Lewis [36] suggests a complex array of factors that determine relevance, ranging from salience to the speaker to practical stakes. In contrast, Dretske [15] is somewhat non-committal, but indicates some commitment to the idea that relevance is a purely *objective matter*, independent of the agent's state of mind.

3.5 Alternatives

What is it for a 'circumstance' to be an alternative to a proposition? What sort of thing is an alternative?

Again, we deliberately use the term 'circumstance' here in a neutral manner. In our minimal model, we have effectively treated alternatives as *unstructured propositions* i.e. as a set of possible worlds. Though, again, we emphasize that in the minimal model, one need not read too much into this choice of terminology: the "worlds" are simply points of evaluation. At any rate, to treat alternatives as propositions is a typical move in the literature, as is the following definition: A is an alternative to P just in case A entails the negation of P.

But this is not the only option. An alternative could be modeled as a *situation* or *set of situations* [3], where formally a situation is akin to a possible world, only that a valuation on situations can be partial. Along similar lines, an alternative could be understood as a *structured proposition*. Other options include: as a centered proposition; as an interpreted sentence; and perhaps others. The choice here might well require more than simply adding fine structure to minimal RA models, but call for perhaps more radical variations on the minimal models and their semantics. For the purposes of this essay, we continue to treat alternatives as unstructured propositions - but this choice is more for convenience than principle.

One further option as to what sort of thing we might take an 'alternative' to be is worth focusing on momentarily: instead of thinking of an alternative as a proposition (of some sort), we could instead think of alternatives as *possible worlds*. Namely, world w is an alternative to proposition P just in case $\neg P$ holds at w. Indeed, there is, in my opinion, a great deal of *ambiguity*, between thinking of alternatives as propositions or as worlds, in some of the key philosophical texts in the RA literature [15, 36].

We can, however, capture the stipulation that alternatives are worlds within our current framework with an appropriate restriction: that only *singleton sets* can occur as relevant propositions. With such a stipulation, the truth clause for $K\varphi$ essentially amounts to: $K\varphi$ holds at world w just in case every relevant world u in which $\neg\varphi$ holds is not a member of the agent's evidence set i.e. is incompatible with the agent's evidence.

3.6 Ruling Out

What is it for an agent's evidence to *rule out* an alternative?

In our minimal models, we captured a notion of a proposition P being *incompatible* with the agent's evidence E (E itself understood as proposition): namely, that the intersection of P and E is empty i.e. at no world is the evidence true and yet P is false. This notion of incompatibility serves as one natural and basic attempt at

capturing the idea of ruling out an alternative, roughly capturing Dretske's idea of having a reason that is *conclusive* with respect to P[9].

This notion of incompatibility may be substantiated in different ways, however. For one thing, the matter as to how exactly to interpret the set E_w - and whether a set of worlds is at all the best modeling device for this purpose - is a delicate issue (briefly touched upon in our discussion of the minimal semantics in section 3.2).

It might be thought, however, that mere incompatibility with the agent's evidence, while clearly *sufficient* for ruling out a proposition, could stand to be supplemented with a richer construal of 'ruling out'. I will emphasize what seem to me two important ways of developing this suggestion. One route is to connect the notion of 'ruling out' with the agent's rational belief: an alternative A is ruled out for the agent just in case it is rational for the agent to find A implausible. Call this the *soft approach* to ruling out. Since the formal treatment of rational belief is itself quite well developed (see, for instance, [4, Ch.7]), tools for the integration of this approach into our formal model are available.

A second approach is to give 'ruling out' a stronger reading: for A to be ruled out for the agent is for the agent to *know* that A is false. Call this the *hard approach* to ruling out. This account has an interesting consequence: if A being an alternative to P means that P entails $\neg A$, then, on the current view, the RA slogan is best understood as commitment to the idea that an agent can know P without knowing all of its entailments. This, then, amounts to a commitment to the denial of knowledge under entailment, even for our cognitively ideal agents, a constraint that can be captured in logical terms.

Thus, the choice between the soft approach and hard approach is intimately tied to the debate concerning the status of epistemic closure principles.

3.7 The Primitive Objects of Relevance

We now consider some less obvious choice points for the RA theorist.

We have been speaking about the relevance of *alternatives* (understood here as propositions). But certain developments in the literature indicate that it is worth considering the relevance of alternatives as *derived* from the relevance of some more fundamental kind of object to which relevance applies.

Heller[10], for instance, suggests that it is *possible worlds* that are the primitive objects of relevance [25, 26]: in a context, some worlds are *similar enough* to the actual world to be considered relevant. Call this the *worlds-first approach*. How then do we recover the relevance of propositions? As follows: A is relevant just in case A holds at some relevant world.

[9] It also goes some way towards capturing Lewis' notion of ruling out (Lewis 1996): for him, A is ruled out just in case it holds at no possible world in which the agent has the same memories and sensory experience.

[10] We may want to place Dretske and Nozick in this camp too.

Switching to possible worlds as the objects of relevance is a simple formal matter: we could, for instance, no longer treat $\mathbf{R}_w$ as a set of propositions, but instead as a set of worlds. We may then alter the satisfaction clause for knowledge as follows:

$$\mathscr{M}, w \vDash K\varphi \text{ just in case } E_w \cap \mathbf{R}_w \subseteq [[\varphi]]_{\mathscr{M}}$$

On the other hand, Jonathan Schaffer has, in a series of recent papers [43, 44, 46, 47], proposed that knowledge claims can only be evaluated relative to a background *question*. Call this the *question-first approach*. To know something, according to this idea, is to know it *rather than* other possible answers to the question, while non-answers and presuppositions to the question are simply ignored. The idea is that in answer to the question "is there a zebra in the cage or nothing at all?" one may know that there is a zebra, but in answer to the question "is there a zebra in the enclosure or a painted mule?" one may not know that there is a zebra. From an RA perspective, this suggests a notion of relevance for propositions: A is relevant relative to (relevant) question Q just in case A is an *answer* to Q. Fortunately, there is an ongoing tradition in the semantics literature from which to draw for treating questions formally [20, 21, 5, 9]. For our immediate purposes, we may understand a question Q, in the formal sense, as a set of disjoint propositions, representing the set of (least specific) answers to that question. An answer to the question is then any subset of a member of Q. A *partial* answer is is any *union* of subsets of Q. A presupposition to the question is any proposition that contains every member of Q.

According to another approach, Stephen Yablo has, again in recent work [53], proposed that knowledge claims can only be evaluated relative to a background *subject matter* or *topic (of conversation)*. Call this the *topic-first* approach. On this view, one can know that the enclosure contains a zebra so long as the subject of painted mules is suppressed. From an RA point of view, this suggests a notion of relevance for propositions as follows: A is relevant relative to (relevant) topic T just in case A concerns (only) that relevant subject matter. Again, fortunately, there are formal tools for integrating subject matters into a formal setting [35]: a subject matter, according to Lewis, can be understood as a *partition* on the space of possible worlds, with two worlds sharing a cell just in case they are exactly the same when it comes to any state of affairs concerning that subject matter.

3.8 Contrast

Is the relevance of a claim A a notion that only makes sense relative to another claim, to which A is to be *contrasted*? Let us say that a theory that answers this question in the affirmative takes the *contrast approach*.

Dretske's theory [13, 15] subscribes to the contrast approach, so will serve as a useful illustration. For Dretske, it does not make sense to describe proposition A as relevant or irrelevant independent of a proposition to which it is to be contrasted. Rather, A can only be understood as relevant or not when understood as an *alterna-*

tive to a particular proposition P (in which case, we can settle whether A might be true were P to be false, following Dretske's notion of relevance).

On the other hand, Lewis' theory [36] does not subscribe to the contrast approach. For Lewis, once the context is fixed, a proposition A is uniformly relevant (or not), no matter which proposition P is being evaluated for knowledge.

Subscribing to the contrast approach, perhaps surprisingly, has far-reaching consequences for an RA theory: Holliday [29, 30] shows that subscription to the contrast approach is the source of the closure failures exhibited by Dretske's theory, while resisting the contrast approach is essentially exactly what allows Lewis to preserve closure in his own theory.

For the purposes of formalization we can capture taking the contrast approach as follows: we may define $\mathtt{K}\varphi$ ('proper knowledge') as follows: $\mathtt{K}\varphi ::= [\varphi]K\varphi$ i.e. proper knowledge of φ is understood as 'knowledge' of φ in the wake of an update that relativizes relevance to φ. That is, we incorporate the contrast approach by stipulating that evaluation of a knowledge claim involves an *update* of the relevancy set (cf. [28]).

3.9 Interaction Principles Between Relevance and Knowledge

In terms of logical principles, what relationship should exist between the relevance of a proposition and knowledge of that proposition? Should there be no such logical relationship? Should the relevance of A act as a necessary condition on knowledge of A (that is, should only *relevant* propositions count as candidates for knowledge)? Should the irrelevance of A be sufficient for $\neg A$ to be known, or for A to be *not* known?

In terms of integration into our framework, stipulating that relevance be a necessary condition on knowledge is at least a simple matter: we simply add the condition $\mathscr{M}, w \vDash R\varphi$ to the clause for $K\varphi$.

4 Survey and Classification of Representative RA Theories

We now exhibit a sample of RA theories, making use of the choice points from section 3 to build some interesting (still relatively abstract) theories that relate to recent and important discussions in the epistemology literature. The first three choice points are the most obvious choice points, and also the hardest to get a formal grip on. For interest and convenience, we essentially focus on the last three: the choice of primitive objects of relevance; the choice between adopting and rejecting the contrast approach; and the choice as to whether to treat relevance as a necessary condition on knowledge.

One goal of this section is to simply exhibit the diversity of RA theories. Another is to illustrate the difference that settling certain choice points can make, and the

implications for other choice points. Another is to demonstrate the formalization of RA theory in action, and demonstrate how formalization can substantively sharpen and otherwise contribute to the philosophical debate.

To achieve this last end, it will be of interest to consider three principles that we can express in our language, and that have bearing on the philosophical evaluation of an RA theory:

- $K\varphi \to K\psi$ whenever $\mathscr{M} \vDash \varphi \to \psi$ **(Closure under entailment)**
- $K\varphi \wedge K(\varphi \to \psi) \to K\psi$ **(Closure under known implication)**
- $K(\varphi \wedge \psi) \to K\varphi \wedge K\psi$ **(Conjunctive distribution)**

By $\mathscr{M} \vDash \varphi$ we mean the standard thing: that φ is true at every world in $\mathscr{M}$.

What is notable about the first two principles is that their validity is *philosophically controversial* (and so where a theory lands on the validity of these principles has philosophical significance) [38]. Recall Dretske's famous example and diagnosis: one may know that the animal in the enclosure is a zebra, without knowing that it is not a painted mule, even though it being a zebra entails that it is not a painted mule. Undoubtedly, Dretske has zeroed in on an important feature of our intuitive judgements. Yet, on the other hand, there are reasons to *resist* dropping closure under known implication: the validity of this principle, it might be said, represents the fact that deductive reasoning from known claims is always a source of knowledge, at least given the idealizations we are working with [32][11].

What is notable about the third principle above is that it is *not* controversial [32, 53]. That a theory invalidates conjunctive distribution may therefore be understood as an unequivocal *strike* against that theory.

4.1 Examples of the Question-first Approach

Let us briefly explore some variations on the question-first approach. Schaffer's work, again, is not explicitly located within the RA tradition[12], but there is nothing stopping an RA theorist from viewing it - or at least certain borrowed aspects of it - in this light. I make no claim in what follows to be representing the details of Schaffer's work entirely accurately (for that, I direct the reader to [43, 44, 46, 47]). In the name of convenience, our aim is to operate according to its spirit, not its letter.

Again, we formally understand a question Q as a disjoint (but not necessarily exhaustive) set of propositions, representing the least specific distinct answers to that question. We integrate this into our RA model as follows:

[11] For recall that we self-consciously model the knowledge of an ideal agent that is always able to "put two and two together" and can therefore maximally extend her knowledge by way of reasoning. To deny of such an agent that closure under known implication is valid is to deny that we ordinary agents are always in principle able to extend our knowledge using self-conscious deductive reasoning by way of known implications.

[12] Though perusal of, for instance, [44, 45] quickly reveals the close ties between Schaffer's views and the RA approach.

Definition 3 (Interrogative RA model). An *interrogative RA model* is a tuple

$$\langle W, \{\mathbf{Q}_w\}_{w\in W}, \{E_w\}_{w\in W}, V\rangle$$

where,

- Every element is as in a minimal RA model, except:
- $\mathbf{Q}_w$ is a set of disjoint propositions.

We first present an RA theory based on these models that rejects the contrast approach (we shall say it is *contrast free*) and does not treat relevance as a necessary condition on knowledge. To do so, we do not need to alter the semantic clauses for that theory: we simply need to define the set $\mathbf{R}_w$ of relevant propositions. First: define $\mathbf{Q}_w^+$ - the set of all answers to $\mathbf{Q}_w$ - as follows

$$\mathbf{Q}_w^+ = \{A \subseteq W \mid A \subseteq A' \in \mathbf{Q}_w\}$$

Then:

$$\mathbf{R}_w ::= \{P \subseteq W \mid P = \bigcup \mathbf{A} \text{ for some } \mathbf{A} \subseteq \mathbf{Q}_w^+\}$$

That is, a proposition is relevant just in case it is an answer (or a partial answer) to the question $\mathbf{Q}_w$.

Effectively, our semantics for $K\varphi$, as detailed in section 3.2, then turns out as follows: $K\varphi$ holds just in case φ holds throughout the partial answers to question $\mathbf{Q}_w$ that are not incompatible with the agent's evidence.

The following result is straightforward to prove. We leave the proof as an exercise for the reader.

Proposition 1. *The above RA theory*

- *validates closure under entailment;*
- *validates closure under known implication;*
- *(and therefore) validates conjunctive distribution.*

By virtue of the validity of closure under entailment, we may therefore note that the current RA theory cannot be one such that ruling out is understood as knowing the negation. So our RA theory is constrained to follow the *soft approach* to ruling out.

Let us now try a variation on the question-first approach: we leave consideration of the contrast approach to another time, but add to our theory that relevance is a necessary condition on knowledge. In the current context this says: one can only know P in response to question Q if P is in fact a (partial) answer to Q. Presuppositions to Q and other non-answers cannot be known - not as a matter of insufficient evidence perhaps, but since these do not *qualify as candidates* for knowledge in the context of the question at issue.

We therefore alter our semantics as follows:

$$\mathscr{M}, w \vDash K\varphi \text{ just in case } \mathscr{M}, w \vDash R\varphi \text{ and } \{A \in \mathbf{U}_w \mid A \subseteq [[\neg\varphi]]_{\mathscr{M}}\} = \emptyset$$

Proposition 2. *Our latest RA theory*

- *invalidates closure under entailment;*
- *invalidates closure under known implication;*
- *invalidates conjunctive distribution.*

One persuaded of the wisdom of rejecting closure will find our altered theory more amenable. But at a cost: conjunctive distribution is lost.

4.2 Examples of the Topic-first Approach

Let us now consider some versions of the topic-first approach. Again, we take inspiration from the work of Yablo, without here attempting to capture the intricate details of his full position (cf. [53])[13].

Following Lewis [35], we understand a topic T as a partition on the set of worlds W. The general idea is this: a topic amounts to a collection of *ways a world could be with respect to that subject matter*, providing an equivalence relation between worlds (two worlds are equivalent just in case they are indistinguishable with respect to how things are with respect to the topic in question) [14]. For instance, if the topic is the 17th century (Lewis' example), then two worlds reside in the same cell of the partition associated with this subject matter just in case affairs with respect to the 17th century are identical in those two worlds.

In the setting of a propositional logic, it is convenient and somewhat natural to instead define a topic T as a set of *interpreted atomic proposition letters* (cf. the semantics of *relatedness logic* [17, 6]). This then serves to *define* a partition on the space W: two worlds w and w' are equivalent just in case they are agree on the truth value of each atom in T. Call this partition π_T.

It is worth remarking on the nature of the partition that a subject matter invokes. We may think of this partition (if non-trivial) as imposing a coarser *resolution* on the space of possible worlds, whereby two possible worlds are treated as indistinguishable unless they differ with respect to the state of the subject matter in question. A subject matter, then, may be understood as controlling the distinctions that are recognized in the space of possibilities: distinctions involving the subject matter are visible, while those that 'cut across' the subject matter are invisible.

Definition 4 (Topical RA model). A *topical RA model* is a tuple

$$\langle W, \{\mathbf{T}_w\}_{w\in W}, \{E_w\}_{w\in W}, V\rangle$$

[13] We mention in footnotes some divergences from important details of Yablo's theory as we proceed.

[14] Yablo in fact embraces a more general conception of a topic: for him, a topic can be associated with a *reflexive, symmetric* relation on the space of worlds, as opposed to an equivalence relation.

where,

- Every element is as in a minimal RA model, except:
- $\mathbf{T}_w$ is a set of atoms, for each $w \in W$.

Let us again begin with an RA theory that is contrast free and does not impose relevance as a necessary condition on knowledge. We may then define the set of relevant propositions as follows:

$$\mathbf{R}_w ::= \{P \subseteq W \mid P = \bigcup \mathbf{C} \text{ for some } \mathbf{C} \subseteq \pi_{\mathbf{T}_w}\}$$

That is: a relevant proposition is identical to a union of cells in the partition determined by the given topic. Other propositions are irrelevant, as they involve distinctions that are 'invisible' to the given subject matter.

We can think of the information the agent's evidence E delivers *with respect to the current subject matter* T as amounting to the smallest union of cells in π_T that contains E. Label this union E^T. Now: given topical RA model $\mathscr{M}$ and world w, define the set $\mathbf{U}_w$ as usual:

$$\mathbf{U}_w = \{A \subseteq W \mid A \in \mathbf{R}_w \text{ and } A \cap E_w \neq \emptyset\}$$

Our semantics are also as before. The net result: $K\varphi$ holds just in case φ holds throughout every cell of $\pi_{\mathbf{T}_w}$ contained in the topic-relevant information $E_w^{\mathbf{T}_w}$. It may once again be checked that the following hold:

Proposition 3. *The above RA theory*

- *validates closure under entailment;*
- *validates closure under known implication;*
- *(and therefore) validates conjunctive distribution.*

It might strike the reader that our last RA theory does not have much appeal, potentially pleasant formal features aside: since $E_w \subseteq E_w^{\mathbf{T}_w}$, there is little epistemic advantage for the ideal agent that adopts a restricted subject matter over consideration of the whole of logical space (for - perhaps intuitively! - refining the relevant topic on this theory tends to *improve* the informational situation of the agent, as this allows her to discern distinctions that were previously ignored). So let us consider one last topic-first theory, one that moves a little closer to Yablo's own presentation: namely, a topic-first approach that both embraces the contrast approach and offers a more nuanced view as to when a proposition is incompatible with the agent's information. That is, in the following theory, relevant subject matter - and so the relevance of propositions - is fixed relative to whatever proposition is being evaluated for knowledge. To accomplish this, we need to provide a fleshed out semantics for $[\varphi]\psi$ expressions.

The idea will be as follows: every sentence φ in the language embodies a natural subject matter: the set T^φ of atoms that occur in φ[15]. Our update operator $[\varphi]$ will

[15] Here we see another divergence from Yablo. For Yablo, the subject matter associated with φ is the set of *ways that* φ *could be true and the ways it could be false*. More precisely: it is the set of

simply update current model $\mathscr{M}$ so that $\mathbf{T}_w$ is replaced with T^{φ}, giving model $\mathscr{M} *_w \varphi$. The semantic clause is as follows:

$$\mathscr{M}, w \vDash [\varphi]\psi \text{ just in case } \mathscr{M} *_w \varphi, w \vDash \psi$$

We now consider a more nuanced account as to when a proposition is incompatible with the agent's information.

Definition 5 (Ordered topical RA model). An *ordered topical RA model* is a tuple

$$\langle W, \{\leq_w\}_{w\in W}, \{\mathbf{T}_w\}_{w\in W}, \{E_w\}_{w\in W}, V\rangle$$

where,

- Every element is as in a topical RA model, except:
- $\leq_w$ is a total order on W, with w a minimal element in the ordering.

Think of $\leq_w$ as a measure of *distance from world w* on the worlds W. We will make use of this ordering to capture a notion as to when evidence E is a conclusive reason for rejecting P: namely, this is the case exactly when in the *nearest worlds* to actuality in which P is true, E is false (cf. [14]). Then, we deploy the following idea: P is incompatible with E just in case E is a conclusive reason for rejecting P[16].

Given proposition A and world $u \in A \subseteq W$, we say that u is $\leq_w$*-minimal with respect to* A just in case there is no world in A closer to w than u, according to $\leq_w$. With this in mind, given ordered topical model $\mathscr{M}$ and world w, define the set $\mathbf{U}_w$ as follows:

$$\mathbf{U}_w = \{A \subseteq W \mid A \in \mathbf{R}_w \text{ and } \exists u \in A \text{ s.t. } u \text{ is } \leq_w\text{-minimal w.r.t } A \text{ and } u \in E_w\}$$

That is: a relevant alternative A to P is now understood to be eliminated by evidence E just in case E is a conclusive reason for rejecting A. A is uneliminated just in case A is relevant and E holds at some nearest A-world to actuality.

Otherwise, our semantic clauses are unchanged. However, we now define 'proper knowledge' of φ as $\mathtt{K}\varphi ::= [\varphi]K\varphi$ and relative relevance as $\mathtt{R}(\varphi,\psi) ::= [\varphi]R\psi$.

The net effect: $\mathtt{K}\varphi$ holds at w just in case for every cell C in $\pi_{T_w^{\varphi}}$ throughout which φ is false, E is a conclusive reason to reject C.

Though the complexity of our models is piling up, the following is still relatively easy to check (see [22] for a comprehensive discussion):

Proposition 4. *Our latest RA theory (where we now understand the following principles in terms of* K *instead of K):*

(what Yablo calls) the *minimal partial models* that succeed in either making φ true or φ false. The reader interested in a proper explication of these notions is directed to [53].

[16] We depart from Yablo here as follows: for Yablo, elimination of alternatives is inspired by the "tracking" account of Nozick: A is eliminated just in case the agent believes $\neg A$ and were A to be the case, the agent would *not* believe $\neg A$. Despite this change in perspective, the technical details for Yablo's account and our own are similar in many respects.

- *invalidates closure under entailment;*
- *invalidates closure under known implication;*
- *validates conjunctive distribution.*

We remark that a key counter-example to closure on the current account is provided by principles of the form: $\mathtt{K}(p) \wedge \mathtt{K}(p \rightarrow p \vee q) \rightarrow \mathtt{K}(p \vee q)$.

This result in fact makes good on one of Yablo's leading motivations for considering knowledge relative to subject matter: the preservation of conjunctive distribution (as inference from a conjunction to a conjunct introduces no new subject matter) while discarding closure (since, according to Yablo, *disjunction introduction* can introduce new subject matter, and so the conclusion should be more elusive than the premises).

5 Conclusion

That concludes our whirlwind tour of the landscape of RA theories. We have accomplished the following: we have seen a number of informal philosophical motivations for embracing the RA approach, ranging from appeal to ordinary linguistic data to the drawing of lessons from famous philosophical examples; we have discussed a minimal framework for formalizing RA theory, and have considered at length various choice points that an RA theorist must decide upon in the construction of her theory; finally, we exhibited four RA theories, by way of setting some of the parameters from the previous section. At the same time, we drew important recent discussions of question and topic-relative knowledge into the RA fold, and demonstrated how the precision of logical techniques can be brought to bear on substantive issues of philosophical evaluation.

Acknowledgements I wish to thank Johan van Benthem and Krista Lawlor for much useful discussion during the development of material in this paper. Thanks for Chenwei Shi and an anonymous referee for many helpful comments based on a close reading of an earlier draft of this paper.

References

1. ANDERSON, A.R., BELNAP, N.D., *Entailment: The Logic of Relevance and Necessity*, Volume I, Princeton University Press (1975)
2. AUSTIN, J. L., "Other Minds", *Aristotelian Society Supplementary* Volume 20, pp. 122 - 197 (1946)
3. BARWISE, J., PERRY, J., *Situations and Attitudes*, MIT Press (1983)
4. VAN BENTHEM, J. F. A. K., *Logical Dynamics of Information and Interaction*, Cambridge University Press (2011)
5. BELNAP, N.D., STEEL, T. B., *The Logic of Questions and Answers*, Yale University Press (1976)
6. BURGESS, J.P., *Philosophical Logic*, Princeton University Press (2009)

7. CARNAP, R., *Logical Foundations of Probability*, University of Chicago Press, (1950)
8. CHELLAS, B. F., *Modal Logic: An Introduction*, Cambridge University Press (1980)
9. CIARDELLI, I., GROENENDIJK, J., ROELOFSEN, F. "On the Semantics and Logic of Declaratives and Interrogatives", *Synthese*, pp. 1-40 (2013)
10. COHEN, S., "How to Be a Fallibilist", *Philosophical Perspectives* 2, pp. 91-123 (1988)
11. COHEN, S., "Contextualism, Skepticism and Reasons", *Noûs* Vol. 33, Supplement: Philosophical Perspectives, 13, Epistemology, pp. 57-89 (1999)
12. DEROSE, K., "Solving the Skeptical Problem", *Philosophical Review*, 104 (1), pp. 1-52 (1995)
13. DRETSKE, F. I., "Epistemic Operators", *Journal of Philosophy* 67 (24), pp. 1007-1023 (1970)
14. DRETSKE, F. I., "Conclusive Reasons", *Australasian Journal of Philosophy* 49, (1), pp.1-22 (1971)
15. DRETSKE, F. I., "The Pragmatic Dimension of Knowledge", *Philosophical Studies* 40 (3), pp. 363–378 (1981)
16. EARMAN, J., *Bayes Or Bust?: A Critical Examination of Bayesian Confirmation Theory*, MIT Press (1992)
17. EPSTEIN, R. L., *The Semantic Foundations of Logic*, Oxford University Press (1994)
18. FLORIDI, L., "Understanding Epistemic Relevance", *Erkenntnis* 69, pp.69-92, (2008)
19. GOLDMAN, A., "Discrimination and Perceptual Knowledge", *Journal of Philosophy* 73, pp. 771-791 (1976)
20. HAMBLIN, C. L., "Questions", *Australasian Journal of Philosophy* 36 (3), pp. 159-168 (1958)
21. HAMBLIN, C. L., "Questions in Montague English", *Foundations of Language* 10 (1), pp. 41-53 (1973)
22. HAWKE, P., "Questions, Topics and Restricted Closure" (Under Review)
23. HAWKE, P., "Truth, Belief and Relevant Alternatives", manuscript (2014)
24. HAWKE, P., "The Problem of Epistemic Relevance", Ph.D dissertation, Stanford University (2016)
25. HELLER, M., "Relevant Alternatives", *Philosophical Studies* 55 (1), pp. 23-40, (1989)
26. HELLER, M., "Relevant Alternatives and Closure", Australasian Journal of Philosophy 77 (2), pp. 196-208 (1999)
27. HINTIKKA, J., *Inquiry as Inquiry: A Logic of Scientific Discovery*, Springer (1999)
28. HOLLIDAY, W. H., "Epistemic Logic, Relevant Alternatives and the Dynamics of Context", in *New Directions in Logic, Language and Computation*, Lecture Notes in Computer Science Volume 7415, pp 109-129 (2012)
29. HOLLIDAY, W. H., "Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism", *Journal of Philosophical Logic*, published online (2014)
30. HOLLIDAY, W. H., "Fallibilism and Multiple Paths to Knowledge", in *Oxford Studies in Epistemology, Vol. 5*, eds. Gendler and Hawthorne, Oxford University Press, forthcoming (2015)
31. KEYNES, J. M., *A Treatise on Probability*, Macmillan, London (1921)
32. KRIPKE, S., "Nozick on Knowledge", in *Philosophical Troubles: Collected Papers, Volume 1*, Oxford University Press (2011)
33. KUHN, T.S., *The Structure of Scientific Revolutions*, 2nd ed. (1st ed. published 1962), The University of Chicago Press (1970)
34. LAWLOR, K., *Assurance: An Austinian View of Knowledge and Knowledge Claims*, Oxford University Press (2013)
35. LEWIS, D., "Relevant Implication", *Theoria* 54 (3), pp. 161-174 (1988)
36. LEWIS, D., "Elusive Knowledge", *Australasian Journal of Philosophy* 74 (4), pp. 549-567 (1996)
37. LUPER, S. "The Epistemic Predicament: Knowledge, Nozickian Track, and Skepticism", *The Australasian Journal of Philosophy* 62, pp. 26-50 (1984)
38. LUPER, S., "The Epistemic Closure Principle", http://plato.stanford.edu/entries/closure-epistemic/, substantive revision Wednesday 4 August 2010 (2001)
39. MACFARLANE, J., "The Assessment Sensitivity of Knowledge Attributions", in *Oxford Studies in Epistemology*, eds. Gendler and Hawthorne, Oxford University Press, pp. 197–234 (2005)

40. MARES, E., "Relevance Logic", http://plato.stanford.edu/entries/logic-relevance/, substantive revision Mon Mar 26, 2012 (1998)
41. NOZICK, R., *Philosophical Explanations*, Harvard University Press (1981)
42. PRITCHARD, D., "Epistemological Disjunctivism", Oxford University Press (2012)
43. SCHAFFER, J., "From Contextualism to Contrastivism", *Philosophical Studies* 119.1, pp. 73-103, (2004)
44. SCHAFFER, J., "Contrastive Knowledge", in *Oxford Studies in Epistemology 1*, eds. Gendler and Hawthorne , pp. 235-271, Oxford University Press (2005)
45. SCHAFFER, J., "What Shifts? Thresholds, Standards, or Alternatives?", in *Contextualism in Philosophy*, eds. Preyer and Peter, pp. 115-130, Oxford University Press (2005)
46. SCHAFFER, J., "Closure, Contrast, and Answer", *Philosophical Studies* 133.2, pp. 233-255, (2007)
47. SCHAFFER, J., "Knowing the Answer", *Philosophy and Phenomenological Research* 75.2, pp. 383-403 (2007)
48. STANFORD, P. K., "Underdetermination of Scientific Theory", http://plato.stanford.edu/entries/scientific-underdetermination/, substantive revision Mon Sep 16, 2013 (2009)
49. STANLEY, J., *Knowledge and Practical Interests*, Oxford University Press (2005)
50. UNGER, P., *Ignorance: A Case for Skepticism*, Oxford University Press (1975)
51. VOGEL, J., "Are There Counterexamples to the Closure Principle?", in *Doubting*, eds. Roth and Ross, pp.13-27, Kluwer Academic Publishers (1990)
52. VOGEL, J., "The New Relevant Alternatives Theory", *Philosophical Perspectives* 13, pp. 155-180 (1999)
53. YABLO, S., *Aboutness*, Princeton University Press (2014)
54. ZAGZEBSKI, L., "The Inescapability of Gettier Problems", *The Philosophical Quartely* 44 No. 174, pp. 65-73 (1994)