Artificial Intelligence 49 (1991) 199–242
Elsevier

Existence assumptions in knowledge representation

Graeme Hirst

Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4

Received November 1989
Revised November 1990

Abstract

Hirst, G., Existence assumptions in knowledge representation, Artificial Intelligence 49 (1991) 199–242.

If knowledge representation formalisms are to be suitable for semantic interpretation of natural language, they must be more adept with representations of existence and nonexistence than they presently are. Quantifiers must sometimes scope over nonexistent entities. I review the philosophical background, including Anselm and Kant, and exhibit some ontological problems that natural language sentences pose for knowledge representation. The paraphrase methods of Russell and Quine are unable to deal with many of the problems. Unfortunately, the shortcomings of the Russell–Quine ontology are reflected in most current knowledge representation formalisms in AI. Several alternatives are considered, including some intensional formalisms and the work of Hobbs, but all have problems. Free logics and possible worlds don't help either. But useful insights are found in the Meinongian theory of Parsons, in which a distinction between *nuclear* and *extranuclear* kinds of predicates is made and used to define a universe over which quantification scopes. If this is combined with a naive ontology, with about eight distinct kinds of existence, a better approach to the representation of nonexistence can be developed within Hobbs' basic formalism.

1. Introduction

Most contemporary logics implicitly or explicitly base the semantics of the quantifiers $\exists$ and $\forall$ on the widely-held ontological assumptions of Russell [66,67] and Quine [56]. A small but growing number of philosophers (e.g., Parsons [50], Routley [65], Lambert [32]) believe that these assumptions are mistaken,[1] and have proposed various alternatives. In this paper, I will

[1] Introducing his work, Parsons says of the Russell–Quine position that "clear progress is rare in philosophy, and I was pleased to have [it as] an example to cite. But as I thought about it more, I became increasingly dissatisfied" [50, p. xii].

discuss the consequences of the Russell–Quine assumptions for knowledge representation formalisms, and show that an adequate treatment requires a multi-faceted view of existence.

My motivation comes from the knowledge representation needs of natural language understanding. As I have argued elsewhere [20], a knowledge representation (KR) formalism to be used in a natural language understanding system for unrestricted text must have (at least) the expressive power of natural language, for otherwise it could not be a target language for semantic interpretation. Moreover, natural languages reflect genuine properties of the real world (with different languages possibly highlighting different properties or viewpoints). Thus, AI research may include exhibiting sentences of natural language and considering how their meaning, and the world it reflects, may be adequately represented—where "adequately" means that the representation permits the same inferences to be drawn as the original sentence. Here, I am concerned with sentences that speak of existence, of nonexistence, or of nonexistent objects.

2. Three ontological slogans

2.1. "Existence is not a predicate"

Immanuel Kant, in his *Critique of Pure Reason* [26, B.625ff], argued that existence is *not* a property that may be predicated of an entity the same way that properties like color and species can be. Kant was responding to an argument by St Anselm of Canterbury [1, Section II] that purported to demonstrate the existence of God *a priori*: his "ontological proof". Anselm's argument was basically this: What we mean by God is, by definition, that entity that is right up the top end of the scale in all desirable properties: the entity that is most wise, most good, and so on. On the scale of existence, clearly actual or necessary existence is better than mere conceptual or possible existence; therefore existence is a defining property of God; therefore God exists.[2] Descartes [10, Section V] later took much the same approach: God has all perfections; existence is a perfection; therefore God exists.[3]

Now, being able to define things into existence like this is metaphysically disturbing, and doesn't really seem possible. Thus, Hume [25, Section IX] tried to show that it is not possible that an entity exist of necessity, and Kant took the position described above, which is often characterized by the

[2]Compare Smullyan's proof [72, pp. 205–206] that unicorns (or anything else you like) exist: To prove that unicorns exist, it suffices to prove the stronger statement that existing unicorns exist. But for existing unicorns to not exist would be a contradiction; therefore existing unicorns exist; therefore unicorns exist.

[3]For the history of the argument, and a discussion of some of the ontological issues mentioned below, see Barnes [2].

slogan "Existence is not a predicate" (cf. Moore [47]). This position is now widely accepted in philosophy [27, p. 160; 52, p. 38]. Nevertheless, while it may have the merit of keeping God off our backs, it does raise difficulties in artificial intelligence.

What I want to show in this paper is that existence *can* be predicated, but (lest God be found to be an emergent property of our knowledge representations; no *deus ex machina* here!) it is neither a single predicate nor a predicate of an ordinary kind.

2.2. "Everything exists"

An adequate treatment of existence in KR formalisms is complicated not only by the problem described above, but also by a related set of difficulties that derive from a position often summarized by the slogan "Everything exists" (cf. Quine [56, p. 1]). That is, there is nothing that doesn't exist, for if it doesn't exist it isn't anything, and statements apparently about nonexistents are either incoherent or can be explained away. The development of this approach is due mainly to Russell [66,67] and, later, Quine [56]. The Russell–Quine position has become so firmly entrenched in twentieth-century Anglo-American philosophy that it is usually accepted without question [50, pp. 1–5]. If we take the slogan literally, then even if existence can be predicated of an entity, it is no more than a tautology; no entities don't exist. And to assert nonexistence of something would be self-contradictory [47,69]. As we will see, this position too is problematic for knowledge representation.

To a large degree, the question seems to be nothing more than what the word *exist* does mean or should mean, and what status is to be assigned to "nonexistent objects". Quine grants two kinds of existence: concrete, physical existence in the world (the kind that Margaret Thatcher has), and abstract, nonphysical existence (the kind that the number 27 has). "Idea[s] in men's heads" [56, p. 2] are included in one or the other of these categories, and so too, I assume, are events and actions. Clearly, this is a wider definition of existence than the kind that Anselm and Descartes's wished to attribute to God. Presumably they intended some divine equivalent of physical existence—able to have causal interaction with the physical world—and would be unhappy with the idea that God existed only in the way the number 27 does. Likewise, Hume and Kant were using the narrower definition when they attacked necessary existence, for many mathematical objects obviously do exist of necessity (the number 27; the least prime greater than 27). So perhaps existence in this other sense, nonphysical existence without causal connection to the world, *could* be a predicate.

2.3. "There are things that don't exist"

Quine's sense of the word *exist* may be wider than Anselm's and Descartes's, but it is still much narrower than that of Meinong [45], who described his position in an oxymoron: "There are objects of which it is true that there are no such objects" [45, translation, p. 83]. For Meinong (like Brentano before him), every thought or idea, such as the idea of a gold mountain, must be "directed toward" some object, and so all objects of thought have being in some sense, even if not real-world existence. Meinong therefore wanted to give status to objects such as the gold mountain, which is not real, and the round square, which is not even possible, arguing that the gold mountain is just as good an object as Mount Everest, and the fact that it is unreal makes no difference. Note that the question is not about the *concept* or *idea* of the gold mountain and whether that exists; clearly, it does. But when we say that the gold mountain is 1000 metres tall, we aren't just talking about an idea; it is not the idea that is 1000 metres tall but the alleged thing that the idea is about.

Russell pointed out that Meinong's approach got into trouble with objects like "the gold mountain that exists"—which isn't real even though existence is part of its definition (cf. footnote 2). It also troubled him that there was any sense in which there can be such contradictory objects as round squares, sets that contain all the sets that don't contain themselves (sometimes known as "Russell sets"), or objects that are not identical to themselves.[4]

Thus the question to be considered is what, exactly, do quantifiers like $\exists$ and $\forall$ quantify over? If an expression begins with "$\exists x$" or "$\forall x$", then what values may be used or considered for x? Do they include Margaret Thatcher, the number 23, World War II, my putting the cat out last night, the late Alan Turing, Sherlock Holmes, the possibility of rain tomorrow, suavity, fear, the set of round squares, the concept of round squares? In other words, what is in the universe of quantification? What exists?

3. What exists?

3.1. What doesn't exist?

The burden on the Russell–Quine position is to explain the apparent counterexamples—to account for the fact that in ordinary, everyday language we can *talk about* certain things without believing that they exist. In this subsection, I will list many examples of reference in natural language to

[4]Parsons [50, pp. 38–42] has argued that a round square is *not* a contradiction in the same way that a nonsquare square is, and that the former is a good object but not the latter. Such distinctions need not concern us in this paper.

seemingly nonexistent entities. My intent is to show that talking about nonexistent objects—and hence representing them, quantifying over them, and reasoning about them—is quite a normal thing to do in natural language. In Section 3.2, I will show how Russell tries to dissolve the problems.

Things that aren't there

Perhaps the simplest apparent counterexample (one that we will see Russell's answer to shortly) is that we can explicitly speak of nonexistence and nonexistent things:

(1) There's no one in the bathroom.

(2) The car I need just doesn't exist [spoken after a long and fruitless search for a suitable car] [76, p. 37].

(3) The perfect chair just doesn't exist.

(4) There's no such thing as the bogeyman; he doesn't exist, and neither does Margaret Thatcher.

(5) Nadia doesn't own a dog.

(6) Round squares are impossible, gold mountains merely unlikely.

We may also speak of events that don't occur and actions that are not taken:

(7) There are no trains to Saginaw on Sundays [i.e., the event of a train going to Saginaw on a Sunday never occurs].

(8) Due to maintenance work on the line, the 6:06 to Saginaw will not run on Sunday.

(9) Today's lecture is cancelled.

(10) The committee's failure to agree on a budget has prevented renovation of the rectory [i.e., the event of the committee agreeing did not occur, and this in turn caused the event of the renovation to not occur].

(11) The workers threatened to hold a strike if their pay claims were not met. The company acceded to the demands, and the strike was averted.

(12) The purpose of the steam-release valve is to prevent an explosion.[5]

[5]In qualitative models of systems (such as boilers) for diagnosis and reasoning, there may be entities such as a steam-release valve whose purpose is "to prevent an explosion". The model does not include or predict any explosion, but the purpose of the valve still has to be somehow accounted for. (I am grateful to Ben Kuipers for this example.)

(13) Nadia refrained from commenting on Ross's new hairstyle.

(14) Ross failed to notice that Nadia had failed to feed the newt.

We can speak of holes, voids, and vacuums; that is, entities that are seemingly constituted by the absence of anything can be spoken of as if they were material objects:

(15) There are too many holes in this cheese [cf. [38]].

(16) Keep your eye on the doughnut, not on the hole.

(17) The pump serves to create a vacuum in the flask.

(18) A complete lack of money led to the downfall of the company.

Existence itself as an object

We can seemingly speak of existence as an object, one that need not exist:

(19) The existence of Pluto was predicted by mathematics and confirmed by observation.

(20) The existence of Vulcan was predicted by mathematics but disproved by observation.

(21) It's a good thing that carnivorous cows don't exist [i.e., the nonexistence of carnivorous cows is a good thing].

(22) A complete lack of money has prevented renovation of the rectory [i.e., the nonexistence of available funds has caused the nonexistence of the renovation].

Claims of reality

We can even (untruly, but not incoherently) assert that unreal objects exist:

(23) I saw a gold mountain near the freeway this morning.

(24) Round squares make me seasick—especially the green ones.

(25) Unreal objects exist.

We can also report such beliefs of others without committing ourselves.

(26) Nadia believes that a unicorn named Old Ironsides has been intercepting her mail and stealing the fashion magazines.

Claims of possibility

We can speak of possible objects and events without committing ourselves either to their reality or unreality, and of objects and events whose existence is merely contingent upon other things.

(27) There may be someone in room 23 who can help you.

(28) If you assemble the parts correctly, you will have created a handsome two-metre model of the CN Tower.

(29) If Ross's mother had accepted that offer of a job in New York and settled down there and married some nice young businessman, she would probably have had a child that would have turned out just like Nadia.

(30) It might rain tomorrow.

Existence at other times

We can refer to things that don't now exist, but did or will. We can speak of things now gone:

(31) Alan Turing was a brilliant mathematician.

(32) Last night's dinner was disastrous.

Sometimes, we may or even *must* use the present tense for things of the past, suggesting that they have some kind of continuing existence:

(33) (a) Alan Turing is a celebrated mathematician [after Barnes [2, p. 48]].
(b) *Alan Turing was a celebrated mathematician[6] [in the sense that he continues to be celebrated].

(34) (a) Alan Turing is dead.
(b) *Alan Turing was dead.

And we can talk of things to come:

(35) Tomorrow's dinner is going to be delicious.

(36) The baby that Diane is planning to have will surely interfere with her violin lessons.

[6] I use the star in the usual way to indicate linguistic ill-formedness.

Fictional and imaginary objects

We can speak of fictional entities and classes as if they really existed.

(37) Dragons don't have fur [52, p. 40].

(38) Sherlock Holmes was the protagonist of many stories by Conan Doyle.

(39) Sherlock Holmes lived in London with his friend, Dr Watson.

And possibly even:

(40) Sherlock Holmes is no longer alive.

Indeed, a large part of the study of literature consists of deriving "facts" about fictional characters that are only implicit in the text:

(41) Holmes regards Dr Watson as a mother figure for whom he has considerable oedipal attraction.

And we can relate fictional objects to objects that do exist:

(42) Nadia models herself upon Sherlock Holmes.

3.2. *The Russell–Quine ontology*

3.2.1. *Paraphrases and the theory of descriptions*

Russell's approach, his *theory of descriptions* [48,66,67], was to regard apparent assertions of existence and nonexistence as merely paraphrases—in logic or a literal English rendering thereof—of other forms in which the assertion is not actually made. Instead, the offending bits are expressed as variables and quantifiers, and the resulting expression is something that can legitimately be true or false. Thus, *Dragons exist* is a paraphrase of *There is at least one thing that is a dragon*:

(43) $\exists x(dragon(x))$.

Since no such x exists, the sentence is false. Likewise, *Dragons don't exist* is a paraphrase of the negation of (43):

(44) $\forall x(\neg dragon(x))$

"For any x, it is not the case that x is a dragon."

Attempts to assert properties of nonexistent objects may be handled in a similar manner:

(45) Dragons like baklava.

$\forall x(dragon(x) \rightarrow likes\text{-}baklava(x))$.

This is vacuously true if there are no dragons [67, p. 229]; but statements about particular dragons would be false:

(46) My dragon likes baklava.

$$\exists x\,(\textit{my-dragon}(x) \wedge \textit{likes-baklava}(x)).$$

This is false because there is no x for which the left-hand side of the conjunction is true. One might instead have used a vacuously true form like that of (45), but the form of (46) reflects Russell's belief that such sentences were false, and also his concerns with definite descriptions (see below).

In the natural language versions of these statements, we have the apparent problem that to even mention dragons seems to give them some sort of existence; to say that *Dragons like baklava* seems to presuppose the existence of the class of dragons. Russell's claim was that on the "correct" reading—the representations above, or literal English glosses of them—the problem dissolves. The logical forms contain no assertion of the existence of a nonempty class of dragons. Moreover, the predicate *dragon* is itself a complex term, and may be regarded as simply an abbreviation for a description such as

(47) $\textit{fire-breathing}(x) \wedge \textit{leather-winged}(x) \wedge \cdots.$

Definite references may also be paraphrased. Thus:

(48) The builder of Waverley station was a Scot.

$$\exists x\,(\textit{built}(\textit{Waverley}, x) \wedge \forall y\,(\textit{built}(\textit{Waverley}, y) \rightarrow y = x) \wedge \textit{Scot}(x)).$$

"One and only one entity built Waverley station, and that one was a Scot." [66, pp. 113–114]

(If the noun phrase being interpreted does not contain sufficient information to uniquely identify the individual, information from context may be added. Thus (48) could also be the representation of the sentence *The builder is a Scot* if the context made it clear that the builder in question was that of Waverley station.) A similar treatment upon *The present king of France is bald* shows the sentence to be false, like (46), because there is no entity denoted by *the present king of France.*[7] Quine [56, p. 7] showed

[7]The problem here is, of course, presupposition failure—the sentence tries to talk about something that doesn't exist, and does so without any of the "redeeming" characteristics of the sentences about nonexistents that were exhibited in Section 3.1. Russell's position on presupposition was famously disputed by Strawson [75], and is no longer generally accepted. Strawson's position was that the presuppositions of a sentence (or, more precisely, of a particular utterance of a sentence) are distinct from its main assertion, and, unlike the main assertion, are unchanged by sentence negation. If a presupposition is false, then the main

how the method can be extended to include proper names, so that sentences about named fictional entities might be paraphrased:

(49) Sherlock Holmes is smart.

$\exists x\,(isHolmes(x) \wedge smart(x))$.

"There is an x that has the property of being Sherlock Holmes, and x has the further property of being smart."

Again, the result is a sentence that is false, for there is no x in the real world that has the property of being Sherlock Holmes.

3.2.2. Problems with the theory

Paraphrasing in this manner immediately disposes of some of the problems mentioned in Section 3.1, but it does so at some cost.

First, all sentences that assert properties of nonexistents are false if specific and true if generic, and negating such sentences doesn't change their truth value! For example, the negation of (46) is:

(50) My dragon doesn't like baklava.

$\exists x\,(my\text{-}dragon(x) \wedge \neg likes\text{-}baklava(x))$.

This is false for the same reason that (46) is. Likewise, the negation of (45), *Dragons don't like baklava*, is true. The underlying problem here, of course, is that English negation and logical negation aren't the same. If we put a "$\neg$" in front of the logical form of (46), we do change its truth value, but that's not what the English word *not* does. In particular, negation in English (and probably in all natural languages) preserves the presuppositions of the original sentence. In the case of (50), alas, it also preserves Russell's erroneous approach to presuppositions (see footnote 7).

A second problem is a technical one in the nature of the paraphrasing task itself: it destroys, quite deliberately, the similarity between the surface form of the sentence and the representation of its meaning. Ryle [69], for instance, regards "quasi-ontological statements" as "systematically misleading expressions"—expressions whose semantic representations, if they have any at all, are quite unlike those suggested by their surface forms. But, as I have argued elsewhere [18,19], there are many virtues in compositional semantic representations in which each element is a direct reflection of a surface constituent of the sentence. While it may not always be possible to

assertion, or the sentence itself, can be neither true nor false; rather, it has no truth value at all. For a review of current approaches to presupposition, see Levinson [34] or Horton [22]. A treatment of presupposition *per se* is beyond the scope of the present paper; for that, see [22,23]. I am concerned here rather with the treatment of the entities that may be felicitously presupposed.

maintain this, the advantages to be gained from it are such that it is not to be given up lightly.

Third, and most seriously, there are, as we saw earlier, sentences about nonexistents for which one's intuition strongly contradicts Russell's theory of descriptions. These include sentences about the defining properties of nonexistents and sentences in which nonexistents seem to have some actual interaction with the real world.

In the first of these classes, we have sentences such as this:

(51) Dragons are small, radish-eating rodents, found mostly in the high Arctic.

$$\forall x\,(dragon(x) \rightarrow (rodent(x) \wedge eats\text{-}radishes(x) \wedge \cdots)).$$

For Russell, this is true, though in any ordinary conversation it would be thought of as false. Likewise, we all agree with Russell and Quine about the falsity of (52):

(52) Sherlock Holmes was stupid.

but we disagree about the reason: in ordinary conversation this sentence is taken as false exactly because its converse is taken as true (cf. Parsons [50, p. 37]).

In the second class are sentences asserting the nonexistence of something. While we might accept representations like (44) for the denial of classes, the denial of the existence of specific entities is trickier. Consider again:

(53) Ross cancelled the lecture.

(54) The (threatened) strike was averted by last-minute negotiations.

On Russell's theory, sentences like these must invariably be false, which is clearly wrong. Notice that paraphrase, in the style of sentence (44), doesn't help here, because these sentences are asserting more than just nonexistence; they are asserting a causal relationship. The expression *The strike was averted* means that the strike never occurred—it did not exist—and that some specific action by someone prevented its occurrence. And which strike was averted? The particular strike that the workers threatened to hold, which has specific properties of time, cause, participants, and so on, that differentiate it from all other real or potential strikes, all properties that could be used when constructing the description in a Russellian paraphrase. But under Russell's view, we cannot truthfully talk about this strike at all, for it does not exist; any sentence that attempts to refer to it will be false. (Note, as before, that we can't get out of this by saying that the reference is to the idea of the strike; it is not the idea that is averted.)

It might be objected that to say *The strike was averted* is a looseness of the English language. One can also use an indefinite reference, and perhaps this is the basic form that should be interpreted: [8]

(55) (When management capitulated,) a strike was averted.

This would yield a representation such as this:

(56) $\exists y(cause(y, \neg\exists x(strike(x))))$.

Someone caused that there be no strike.

(We shall blithely allow *cause* as a predicate that takes a proposition in its second argument and asserts that the entity in the first argument caused the second argument to be true.) The problem with this tack is the need to say exactly what didn't happen. After all, there have been a lot of strikes that weren't averted; but (56) says there were no strikes at all. Clearly, some identification from the context is necessary: what was averted was a strike by some particular set of workers at some particular time over some particular claim—so we must identify *the* strike in context, and we're back to where we started.

Another objection might be that the proper paraphrase is *The strike that was planned was averted*, the claim being that the strike does exist, nonphysically, like mathematical objects, by virtue of its having been planned. (This would explain why it sounds a bit funny to say *The accident was averted* instead of *An accident was averted* (cf. above), as accidents aren't planned.) The problem with this is that we then have to explain what it would mean to avert an abstract object. Perhaps it means averting the physical realization of this nonphysical object—in effect, the instantiation of a concept. This view follows the lines of Frege's argument that existence *is* a predicatable property, but is a property of concepts, not individuals [13, Section 53, p. 65; 15, pp. 18–19,32,76–66]. To say that something exists is to say of a concept that it has an extension. So, for Frege, the error in Anselm's argument was applying the predicate wrongly—applying it to an extension, God, rather than a concept, the concept of God. On this view, the sentence *Dragons exist* would mean that the set of extensions of the concept of dragons is not empty. And to say that the strike was averted would be to say that there was caused to be no extension that corresponds to the concept of the particular strike in question (specified, in the manner of Russellian paraphrase, in sufficient detail to be unique).

This approach has generally been regarded as philosophically unsatisfactory [73, p. 90]. [9] It seems to just sidestep the problem terminologically,

[8] Barry Richards, personal communication.

[9] It is simply ignored, for example, by Moore [47] and Prior [54] in their reviews of the problem, and is peremptorily dismissed by Parsons [50, p. 216].

leaving us no wiser as to the nature of the first-order property that all extensions allegedly have—Frege called it "actuality" (*Wirklichkeit*) [13, Section 26, p. 35] (see also [8, p. 194, n. 7])—which, by any name, is the property that we are interested in here. So the problem can't be reduced to one of concepts.

But perhaps we could say that if the strike was planned, it exists as a "future object". To examine this, we must consider the role of time. Russell provides no treatment of existence at times other than the present, but we can speculate on how he would extend his theory to do so.

Let's consider the simpler case first: the past. It is unclear from Russell's account how he would paraphrase, say, *Alan Turing was smart* and *Alan Turing is dead.* That is, would he allow the scope of quantification to include past entities? Doing so would let the first of these sentences be paraphrased like any other, and the past-tense verb would just be an artifact of the pastness of Alan Turing himself, not included in the paraphrase:

(57) Alan Turing was smart.

$$\exists x\,(isTuring(x) \wedge smart(x)).$$

This would then be a true sentence, unlike *Sherlock Holmes was smart.* But this doesn't work for the second sentence:

(58) Alan Turing is dead.

$$\exists x\,(isTuring(x) \wedge dead(x)).$$

It doesn't work because Turing wasn't dead when he existed, and the verb tense hasn't behaved as in (57). At a minimum, we need to add some notion of time points or intervals such that propositions can be true at some times and not others; thus, (58) would be true today, but false in 1945 and 1862—false in 1945 because Turing was still alive, and false in 1862 because he hadn't then come within the scope of the existential quantifier.

Thus the universe would be seen as travelling through time, collecting up entities into its ontology as it proceeds. Once a thing has started to exist, it never stops. This helps represent sentences (57) and (58), but I don't think this view can be pleasing for the everything-exists gang, for the fact remains that Alan Turing does not now exist in the world any more than the gold mountain does, nor does he seem to exist as a mathematical object. (The idea of Turing continues to exist, but it's not that that's dead.) There doesn't seem to be any good reason why his brief time on earth should give Turing any subsequent ontological advantage over the gold mountain. [10]

[10] A rejoinder that I shall not take very seriously: Alan Turing does in fact still exist, or at least his soul does, in Heaven or Hell or somewhere like that. On this view, one might say that the best paraphrase for *Alan Turing is dead* is one of these:

These problems may be seen even more clearly if we now consider future entities, such as the strike that the faculty are threatening to hold. We can talk about this just as easily as we can about Alan Turing (albeit with less certainty)—it will be long and nasty, it will cause the university president to resign, it may never happen (!). For Quine, certainly (and presumably for Russell—guilt by association), the strike is merely a "possible object", to be kept out of one's ontology at all costs (cf. his arguments against the existence of the "possible man in the doorway" [56]). So now the averted strike is out on two separate counts, each fatal on its own. When it was still a planned strike, it was merely a possible object; after it was averted, it became a past object as well.

But for knowledge representation and natural language understanding, this is simply not acceptable. I have shown above that objects like Alan Turing and the averted strike must be able to be represented, quantified over, and reasoned about just as much as Margaret Thatcher. So the Russell–Quine

(i) Alan Turing's body doesn't exist (or no longer exists).
$\neg\exists x\,(bodyOfTuring(x))$

(ii) Alan Turing is in the afterlife.
$\exists x \exists y\,(isTuring(x) \wedge afterlife(y) \wedge in(x, y))$

Form (i) is undoubtedly true, and the truth of form (ii) depends on whether there is an afterlife and if so who's there, issues that I will not solve in this paper.

The value of this particular rejoinder is to draw attention to the cultural bias in the expression of the problem; perhaps we say that Alan Turing *is* dead just because English reflects our long cultural history of belief in a soul and an afterlife. If we are careful to avoid such bias in our language, we will be able to analyze the problem correctly (or so said a large twentieth-century school of philosophy). Notice, for example, that English offers no analogous expressions for the past existence of objects to which we do not (culturally) attribute an afterlife; if my wristwatch has ceased to be, I can say *My wristwatch was destroyed* but not *My wristwatch is destroyed* (and only as a joke or metaphor, *My wristwatch is dead*). Thus when we say that Turing is dead, our paraphrase should be no more than that there is no x such that $isTuring(x)$; and that this statement was false at an earlier time is an implicature of the word *dead*.

I don't think that this argument goes through. There are too many other things we can say about entities of the past that seem to presume their continued existence:

(iii) Alan Turing {is | *was} a celebrated mathematician.

(iv) Nadia models herself upon Alan Turing.

(v) Nadia knows more about NP-completeness than Alan Turing ever did. [Although Turing is referred to in the past tense, the entity *Alan Turing's knowledge of NP-completeness* is available for comparison with an entity, Nadia's knowledge, that exists in the present and did not exist at the time of Turing.]

(vi) Nadia modelled her new sculpture upon my old wristwatch (which was destroyed last year).

(vii) The Flat Earth Society is now disbanded.

position is inadequate. Unfortunately, as I will show next, most knowledge representation formalisms share the Russell–Quine deficiencies.

4. Existence assumptions in KR formalisms

To what extent are knowledge representation formalisms able to deal adequately with existence and nonexistence? The universe of discourse of a system is, of course, circumscribed by what's in its knowledge base; but given that nonexistent entities may have to be included (and, in a full NLU system, *must* be included), how does the average formalism behave?

For the most part, KR formalisms are Russellian in their approach to ontology. To use a term is to assert that it denotes, and, in particular, that it denotes an extant entity [79]. To assert, for example,

(59) Ross cancelled the lecture.

cancelled(*Ross*, *lecture23*).

implies for most systems (e.g., KRYPTON [3] and Sowa's conceptual graphs [74]) that *lecture23* exists just as much as *Ross* does, even if the expression says that it doesn't.

4.1. Platonic-universe approaches

Not all KR formalisms impute existence to denotations of their terms. A simple first-order system in which (ignoring all the philosophical wisdom discussed above) existence is a predicate like any other has been proposed by Hobbs [21] in his paper entitled "Ontological promiscuity". The "promiscuity" of the title refers to the Meinong-like inclusion of nonexistent objects, including the reification of events and properties as objects; [11] Hobbs' set of objects is a Platonic universe, "highly constrained by the way the ... material world is" (p. 63). The quantifiers $\exists$ and $\forall$ range over this universe, and all variables are assumed to denote some entity in it. In general, the formalism is deliberately simple and "flat", without modals, intensions, or even negation. (Hobbs' aim in the paper is to show that predicates in his system suffice instead.)

In this approach, no object mentioned in a representation is assumed to exist in the real world unless such existence is either explicitly stated

[11] Treating events as objects, in the style of Davidson [9], is a position that I have adopted in this paper and assumed to be relatively uncontroversial even for supporters of the Quine–Russell position. Treating properties as objects is a separate question somewhat orthogonal to the concerns of the present paper; suffice it to say here that Quine and Russell would not, I think, approve.

or axiomatically derivable. For example, *Ross worships Zeus* is represented as:

(60) $worship'(E, Ross, Zeus) \wedge Exist(E)$.

The first conjunct says that E is a worshipping by Ross of Zeus, and the second says that E exists in the real world. (Do not confuse the predicate *Exist*, which denotes real-world existence, with the quantifier $\exists$, which ranges over the entire Platonic universe.) The predicate *worship'* is existentially transparent in its second argument but not its third. This means that the real-world existence of E implies the existence of Ross but not that of Zeus. That is, it is an axiom of the system that:

(61) $\forall E \forall x \forall y((worship'(E, x, y) \wedge Exist(E)) \rightarrow Exist(x))$

Hobbs shows that with an adaptation of Zalta's system of abstract objects [80], this approach is able to deal with several problems of opaque contexts that are usually thought to require higher-order representations, while at the same time remaining (moderately) faithful to the surface form of the English sentence.

Although Hobbs mentions nonexistence only briefly, it is clear that by extending his approach we can account for some of the problems mentioned above. Just as transparent argument positions entail existence, we will allow an argument position to be *anti-transparent*, entailing that the object in that position does *not* exist. (Anti-transparent positions are not to be confused with Hobbs' opaque positions, which entail nothing.) We can then represent the prevention of the occurrence of the strike:

(62) The strike was averted.

$strike(s) \wedge \exists x(Exist(E) \wedge avert'(E, x, s))$.

"s is the strike (identified from context), and in the Platonic universe there is an x such that x averted s, and the averting E really exists."

It would be stipulated that *avert'* is transparent in its second argument and anti-transparent in its third—that is, in (62), the existence of the averting, E, would imply the existence of the averter but the nonexistence of the strike:

(63) $\forall E \forall x \forall y((avert'(E, x, y) \wedge Exist(E)) \rightarrow (Exist(x) \wedge not(Exist(y))))$.

The existence of existence also seems representable. Hobbs has a "nominalization operator", $'$, which turns an n-ary predicate into an $(n+1)$-ary predicate whose first argument is the condition that holds when the base predicate is true of the other arguments. We saw this above with ternary predicates such as $worship'(E, Ross, Zeus)$, derived from the binary predicate $worship(Ross, Zeus)$. Since *Exist* is just another predicate, there is nothing to stop us nominalizing it:

(64) The existence of carnivorous cows is predicted by GB theory.

$$Exist'(E_1, carnivorous\text{-}cows) \wedge predict'(E_2, GB\text{-}theory, E_1) \wedge Exist(E_2).$$

"E_1 is the existence of carnivorous cows, E_2 is the prediction of E_1 by GB theory, and E_2 exists (though E_1 might not)."

On the other hand, there is no treatment of fictional objects. Nonexistent objects can be mentioned, as we saw in the assertion of *Ross worships Zeus*, but there is nothing that lets us say that Zeus exists in fiction whereas the Giant Cosmic Groundhog (which I just made up) and the averted strike do not. An obvious move is simply to add a predicate *Fictional* to the formalism. Then *worship'* would have the property that its third argument must exist either in the real world (like Nadia, whom Ross also worships) or in fiction (even if only a small fiction in Ross's mind). Hobbs' Platonic universe would now have a tripartite division into the existent, the fictional, and all the rest. [12]

But there is no reason to stop at a tripartite division. Following Fauconnier [11], we can divide the Platonic universe into many different, overlapping ontological spaces, one for each different work of fiction, each theory or hypothesis, each different perception of the world. In Fauconnier's theory, the "reality" [11, p. 17] of some agent is the top-level universe, and each division, or *mental space*, is a subset of the entities in that reality and the relationships in which they participate. A mental space may include any existent or nonexistent entity that the agent thinks about. Fauconnier shows how mental spaces can serve in a semantic theory for natural language, accounting for such phenomena as embedded belief contexts, presuppositions, and counterfactuals. This is compatible, I think, with Hobbs' approach, and indeed is implicit in the approach that I develop in Section 7.1 below.

But so far, this approach doesn't give an adequate treatment of objects like Alan Turing—we can't talk about Turing's different and divergent statuses at different times. Hobbs' notion of time is based on English verb tenses. An assertion can be said to be true in the past or future. So one could say that Alan Turing's existence is true in the past—but it has to be *all* the past. A better approach is developed in the TELOS system of Koubarakis et al. [28,29], in which the truth of an assertion may be limited to any

[12] I will resist the temptation to be side-tracked onto the question of characterizing more precisely what it means to be fictional rather than just nonexistent; see [78] for discussion.

time interval, and one can quite literally have objects like *Alan Turing 1912–1954*.[13]

In addition, it seems that Anselm's fallacy is valid in the system. Although Hobbs gives no examples of definitions, it seems that *Exist* can be used directly or indirectly as a defining characteristic, since it's just another predicate. Its direct use in a definition could be prohibited by stipulation; but preventing its indirect use is not possible, as it is a deliberate feature of the system that existence can be axiomatically derived from various assertions—one has to be allowed to define predicates with transparent arguments. Thus, following Descartes's version of the fallacy,[14] one could define the predicate *perfect* to be transparent in its (sole) argument, and then assert that God is, by definition, *perfect*:[15]

[13]Hobbs has pointed out (personal communication) that a similar effect could be developed in his system by treating times as entities, and asserting that each particular existence occurs at a particular time.

[14]Anselm's original version, as I glossed it in Section 2.1 above, is second-order and so would not be expressible in Hobbs' system as it presently stands (but see Lewis's first-order possible-world formalization of a slightly different reading of Anselm's argument [36]). Assuming the addition of second-order quantifiers to Hobbs' formalism, we could express Anselm's argument as follows:

(i) $\forall S \forall P(scale(S) \wedge maximum(S, P) \rightarrow P(God))$.

"For any scale S such that P is the property of being at the maximum point on that scale, God has property P; i.e., God is up the top end of the scale in all (desirable) characteristics."

scale(*Wisdom*),
scale(*Lovingness*),
scale(*Existence*).

"Scales include wisdom, lovingness, and existence."

maximum(*Wisdom*, *Omniscience*),
maximum(*Lovingness*, *AllLoving*),
maximum(*Existence*, *Exist*).

"The top end of the wisdom scale is omniscience, of the lovingness scale is being all-loving, of the existence scale is real-world existence." (*Necessary* real-world existence would be an even stronger condition (cf. Section 2.1 above), but Hobbs' standard predicate *Exist* suffices to make the point.)

[15]It might be argued that this is a *virtue* of the system. The system is supposed to represent natural language; we can express Anselm's fallacy in natural language; so the system should be able to represent Anselm's fallacy. This is true; but it doesn't follow that the fallacy should be *valid* in the system; after all, it isn't valid in natural language (but cf. [36]). Just as a formalism should be able to represent entities regardless of their existence, it should be able to represent arguments regardless of their validity—but that goes beyond the scope of this paper.

It might also be suggested that the validity of Anselm's fallacy in the system is nothing more than an example of "garbage in, garbage out". Write some silly axioms and you get a silly answer. One can perform analogous abuses in *any* formalism, such as just directly stating the existence of God (or anything else) as an axiom:

(i) *Exist*(*God*),
Exist(*Giant-Cosmic-Groundhog*).

(65) $\forall x (perfect(x) \rightarrow Omniscient(x)),$
$\forall x (perfect(x) \rightarrow AllLoving(x)),$
$\forall x (perfect(x) \rightarrow Exist(x)).$

"To be perfect is to be omniscient, all-loving, and existent."

$perfect(God).$

"God is perfect."

The same logical cornucopia will produce the perfect armchair, the perfect automobile, and the perfect lover at little extra expense.

4.2. Intensional approaches

Although it was important for Meinong that thoughts and ideas could be directed to nonexistent objects, I have said little up to now, except in passing, about ideas, intensions, and concepts. Indeed, both Russell and Hobbs were at pains to avoid the standard Fregean distinction [14] between intension and extension (*Sinn* and *Bedeutung*). But even Quine grants ideas a place in his universe (see Section 2.2 above); so we now turn to this topic. I will use the terms *concept*, *idea*, and *intension* interchangeably below; the technical differences between them will be unimportant. Likewise, I will conflate *extension* with the *denotation*, *realization*, or *instance* of an idea.

An adequate treatment of concepts as "first-class objects" has often eluded knowledge representation systems. By a first-class object, I mean here an object that can be referred to as an individual in its own right, be used in inference, be a component of other objects, and so on. This would be necessary if we were to act on the suggestion (Section 3.2.2 above) that the sentence *The strike was averted* be represented as the prevention of the realization of an instance of the concept of strikes. Now, because concepts are used to define other objects, many systems accord them a special status that precludes their simultaneously acting as ordinary objects or individuals. A typical example is Charniak's FRAIL [6], a language in which concepts are *generic frames*, but inference can be carried out only on *instances* of those frames; it is not possible for a frame to be simultaneously generic and an instance. In KRYPTON [3], which makes a careful separation of

Only a clumsy stipulation could prevent such deliberate abuse of a formalism, and there seems little reason to bother doing so in any practical use of the system.

The point that this objection misses is that Hobbs' system *encourages* the use of transparency axioms such as (61), and a practical system would have many hundreds of them. Situations like that summarized in (65) might arise from an unexpected interaction of scattered axioms and definitions in the system, each of them individually acceptable and intended to do nothing more than to define various concepts and terms. In this connection, it's also worth noting that one of Frege's motivations in [13] was to prevent spurious mathematical objects being "defined into existence" by ill-formed definitions.

"terminological" knowledge (which goes in its "T-box") and assertions about the world (in its "A-box"), it is possible to reason with the terminological knowledge, which can be thought of as statements about concepts, but concepts *per se* can still not be reified as first-class individuals.

Languages in which concepts *are* first-class objects include McCarthy's first-order language [43,44], Shapiro and colleagues' SNePS [40,71], and Sowa's conceptual graphs [74]. Such languages must provide a mechanism to relate objects to the concepts of which they are instances. For example, Sowa's conceptual graphs tie concepts and their extensions together by notational means. Thus `[CAT:*]` represents the concept of cats, and `[CAT:#234]` represents some particular cat (namely, cat number 234). The notation `[CAT:*x]` represents the individual concept of a cat: a single cat, but not any particular known one; the `x` may be thought of as a variable, so that all occurrences of `[CAT:*x]` must refer to the same (unknown) cat, but `[CAT:*y]` might be a different one. These different types may be used interchangeably (with different meaning, of course) in the graph representations that can be built. However, all graphs are implicitly existentially quantified; that is, the ontology is implicitly Russellian.

The SNePS network formalism is of special interest, as Rapaport [63] has suggested that Parsons' theory (Section 6 below) could give it a formal semantics. In SNePS, *all* entities are intensions, and extensions *per se* are not used. This is because SNePS takes representations to be those of the knowledge of an agent rather than representations of the world directly. The intensions are connected to reality only through the agent's perception. Thus SNePS is free of extensions only for an external observer of the system. The SNePS objects used by a computational agent that employs the formalism (such as Rapaport's CASSIE [71]) are the concepts in that agent's "mind", so to the observer they are intensions. To the agent itself, however, they are subjective extensions, identified with its perceptions of reality. Shapiro and colleagues show only individual concepts, such as the node *John* representing the idea of John;[16] I assume that if the agent is to think about the idea of John, it will need a node that represents the idea of the idea.

McCarthy's first-order language adapts the approach taken by Frege (Section 3.2.2 above). McCarthy includes both concepts and extensions as entities in his language, though, unlike Frege, he does not formally distinguish them from one another.[17] A function called *denot* maps concepts to the entities, if any, that they denote. (Thus individual concepts such as *John* are mapped to an individual, and generic concepts like *Dog*, not explicitly mentioned by McCarthy, would presumably be mapped to an appropriate

[16] For simplicity, I am ignoring Shapiro's careful distinction between nodes and their names.

[17] The typographical distinctions in McCarthy's formulas are for the reader's convenience, and are not part of the theory.

set of individuals.) Following Frege, the predicate *Exists* is true of those concepts for which there is a denotation.[18] Predicates for concepts may be defined "parallel" to those for denotations. For example, if *ishorse* is a predicate true of horses, then *Ishorse* can be defined as a predicate true both (i) of concepts for which *ishorse* is true of their denotations, and (ii) perhaps also of some concepts that don't have denotations, such as *Pegasus*.

Philosophically, this suffers from the same problems as the Fregean approach upon which it is based. As a representation formalism for AI, it has the advantage of simplicity in being first-order, but also the consequent disadvantage that intensions cannot have any special status. Indeed, generally speaking, knowledge representation formalisms that treat concepts as first-class objects do not formally distinguish them from individuals. (Those that don't, do; they have to, in order to discriminate against them.) I don't know of any principled reason for this. Such systems are *weakly intensional* systems, countenancing intensions but not making anything special of them. In contrast, *strongly intensional* systems take intensions to be not just first-class objects but objects of a distinct kind.[19] Montague semantics [46] is a good (noncomputational) example of a strongly intensional system. A strongly intensional system will be surely necessary for an ontologically adequate treatment of intensions. McCarthy could use his *denot* function to map intensions to their extensions, but going in the opposite direction requires an operator, as in Montague semantics. The examples of Section 3.1 show such operations to be frequently necessary, and the modes of existence to be discussed in Section 7.1 below suggest that a diverse set of operators may be required.

The story so far. We want to represent natural language sentences about existence and nonexistence. Philosophers tell us (with some justification) that we'll get into trouble if we construe existence as a predicate. But following this advice leaves us with a KR formalism too weak to do the job. And so far, even formalisms that ignore the advice are inadequate or troubled.

Next. Some suggested solutions.

[18] So McCarthy's predicate is not to be confused with Hobbs' (Section 4.1 above). McCarthy's *Exists* is a predicate true of concepts that have real-world denotations; Hobbs' *Exist* is true of the real-world objects themselves.

[19] This distinction is due to Graeme Ritchie (personal communication).

5. Free logics and possible-world formalisms

5.1. Free logics

Another way that has been suggested around the Russell–Quine problems is the use of *free logics*. A free logic is a logic that makes no assumptions about existence—specifically, a logic that tolerates terms that have no denotation in its universe, never quantifying over them.[20] For example, Woodruff's system **UE** [77] is a free logic with truth-value gaps (i.e., with the truth values **t**, **f**, and **u**) and a distinction between assertions of truth and assertions of nonfalsity. Nondenoting terms have no interpretation at all, and a predicate need only have truth value **t** or **f** if all its arguments denote. Thus the system is explicitly Strawsonian. In contrast, Schock's free logic [70] has only two truth values, and (in the style of Frege) uses the empty set as the "denotation" of nondenoting terms. Both systems have an "existence" predicate, which is true just of those terms that denote. (See [42] for a survey of free logics and their properties.)

At first sight, free logics seem to be an attractive solution in KR to the problems of Russellianism. Free logics are a conceptually easy extension of classical systems; deduction systems already exist for them; and truth-value gaps are already a focus of research in the field (e.g., Patel-Schneider's four-valued logic [51]). From a natural language perspective, free logics help avoid Russellian paraphrases, thereby leading to a more compositional semantics—we can use any object "as is". So if we want to say that Alan Turing was smart, we can say (66) directly, with no need for an existential quantifier:

(66) *smart*(*AlanTuring*).

But alas, free logics turn out to have most of the same problems for natural language understanding as Russell's standard logic. We are allowed to use the term *AlanTuring*, but Alan Turing himself is still not in the universe for quantification. Sentences like (66) need not be false (at least in Woodruff's logic), but (except in a trivial, unhelpful way) they still can't be true.

5.2. Possible worlds and their populations

Clearly, then, the problem is to somehow bring Alan Turing, the averted strike, Sherlock Holmes, and our other nonexistent entities within the set of entities of which true predications may be made, while not allowing them to be considered existent. One suggestion for this is the use of the concept of

[20]Hobbs' system (Section 4.1 above) is not a free logic. While it makes no assumptions about real-world existence, it does assume that all terms denote something in the Platonic universe, and it quantifies over them.

a *possible world*. Then we could say that such entities are members of other possible worlds—including worlds of fiction and worlds of other times—but not members of the actual world. (Notice that possible worlds are themselves nonexistent objects that we can talk about in the real world.)

There are many different versions of the notion of possible worlds, and a complete survey would be beyond the scope of this paper.[21] Generally, however, a world is construed as a maximal, consistent state of affairs. That is, a world is complete, in the sense that nothing is unspecified, and the specifications are not contradictory (e.g., [31, p. 18; 53, p. 44; 78, pp. 103–104]). A world W' is *possible relative to another world* W (or, equivalently, is *accessible from* W) if, intuitively, the state of affairs of W' might possibly have obtained instead of W, where "possibly" may be construed as broadly or narrowly as one likes. For example, one could take it as logical possibility, and so permit, in worlds possible relative to our own, pigs that fly by means of anti-gravity grunting; or one could take it as physical possibility and so require the flying pigs to employ wings; or one might require the worlds to be very similar, and so exclude flying pigs altogether.

Given the notion of possible worlds, the question then arises as to what the individuals that populate the worlds are. On the one hand, we intuitively want to say that, by and large, the same individuals turn up in more than one world, even if they have different properties. So the Margaret Thatcher who won a certain election in the real world is the same individual who lost that same election in a different possible world. Of course, some worlds will have individuals that our world doesn't, such as the baby that Laura had in the world in which the condom broke; and some worlds will lack individuals that ours has, such as Margaret Thatcher in the world in which her parents never met one another. On the other hand, some philosophers, from Leibniz on [53, p. 88], have held that, since individuals in different worlds have different properties (even if only the property of being in some particular world), they must be distinct individuals. David Lewis [35] has proposed that although individuals can be in at most one world, they can have *counterparts* in other worlds—possibly several of them in a single world. The counterparts of an individual are those things, if any, in other worlds that are most similar to that individual. The counterpart relationship is not transitive or symmetric, and can be one-to-many and many-to-one.[22] Taking

[21] Indeed the very notion of a possible world is controversial, as are the quantified modal logics associated with them (see below). The bad guy, once more, is Quine. His objections are given in [57], which is reprinted, with replies from the other side, in [39]. A summary is given by Plantinga in [53].

[22] If we limited counterparts to at most one per world, and made the relationship symmetric and transitive, then we could identify equivalence classes of counterparts with individuals, and the approach would become effectively the same as its competitor; but such limitations are explicitly not Lewis's intent [35, p. 28].

the middle ground between these positions, Chisholm [7] and Purtill [55] have suggested that "small" changes in properties across worlds preserve the identity of an individual, but cumulatively, such changes will eventually lead to it becoming a different individual, even though the transition point may be blurry.

All routes lead to trouble here. If we restrict each individual to a single, independent possible world, we get nowhere with our project in this paper, accounting for the role of nonexistent objects, such as the averted strike, in this (or some) world. Other possible worlds will be quite unconnected to the world under consideration. Lewis's counterpart theory would serve to forge a connection between the worlds, but the theory has many problems (see [31,53]). For example, when we go looking for the averted strike, what are we looking for? An actual strike, presumably, that has no counterpart in the real world. But if it has no counterpart in the real world, how can we identify it? In what sense is it related to the strike that, in the real world, was averted? Presumably, it has the same players, cause, time, location, and so on. But to say that is to reify the averted strike in the real world, and that's exactly what we're trying to *avoid* doing.

But if we allow an individual to turn up in different worlds with different properties, anarchy is not far away. For example, Margaret Thatcher could turn up as a man named István Regoczei who leads a motorcycle gang in Budapest, while Michael Jackson is a woman named Margaret Thatcher who becomes prime minister of the United Kingdom and Ronald Reagan is a palm tree in Florida. [23] This can be prevented if we stipulate that each individual has certain properties, *essences*, that are the same in all worlds; the problem, of course, is in deciding which properties they should be. [24] In most practical AI systems we would want to be quite conservative, and consider the essences to be "important" properties, such as being of a certain natural kind.

5.3. *A naive formalization*

Let's agree, then, that the same individual may occur in many possible worlds, and that we can constrain the accessibility relation between worlds by stipulating essences. Mathematical objects such as numbers will occur in

[23]Note that an individual's having a different name in different worlds is *not* inconsistent with Kripke's notion of a name as a *rigid designator* that picks out the same individual in all worlds [31]. Kripke is quite explicit [31, pp. 49, 62, 77–78] that a name, when used by us in this world, picks out in other worlds the same individual as it does in this world, regardless of that individual's name in the other worlds.

[24]The literature is divided on whether there really are properties that an individual necessarily has in all logically possible worlds; Plantinga [53] and Lewis [35] say yes; Parsons [49,50] is less certain. Regardless, we can always *stipulate* essences as part of our definition of accessibility between the worlds we wish to consider.

all worlds. We can now see whether this approach will provide an adequate representation of sentences about nonexistent objects. We shall start at a rather naive level with standard, first-order logic, and suggest that instead of asserting the nonexistence of, say, the cancelled lecture, we need merely say that its existence is in some other possible world. For a possible world W, let the *domain* of W, written $D(W)$, be the set of individuals in that world. Let R be the real world. Then we have:

(67) Today's lecture is cancelled.

$$\exists W(W \neq R \wedge \exists x \exists y(x \in D(R) \wedge y \in D(W) \wedge y \notin D(R) \wedge \\ \mathit{person}(x) \wedge \mathit{today's\text{-}lecture}(y) \wedge \mathit{cancel}(x,y))).$$

"There exists some world W, not equal to the real world R, among whose individuals there exists a y that isn't among the individuals of R and that is today's lecture, and there exists some individual x in R who is the person who cancelled y."

There are many obvious immediate objections to this. First, it seems to say too much. Someone who asserts *Today's lecture is cancelled* is surely not intending to say anything about the lecture's existence in other possible worlds. So the part of (67) that says $y \in D(W)$ for some $W \neq R$, is just unhelpful, irrelevant baggage. Indeed, it's a tautology, for on the theory that we are trying to apply here, *everything* has that property.[25]

Second, (67) invokes the relation *cancel* between objects that are in different possible worlds. This seems just a little mysterious. How is it possible for a relationship to hold across worlds at all? How exactly was an x in one world able to do something to a y in another? If the cancellation itself is an action in the real world, how can one of its components be in a different world? All these points need clarification.

Third, (67) contains quantification over possible worlds, which are nonexistent objects, and quantification over objects in the domains of possible worlds, some of which exist (because they are also in the real world) and some of which don't. So all of the Russellian problems (Section 3.2.2 above) immediately apply, and the sentence must be false. But this is unfair! The whole point of bringing in possible worlds was so that quantifiers could gain access to the objects in their domains. Clearly, we can't play this game by Russell's rules; we need to consider a system that's more hospitable to possible worlds.

[25] That is, *every* object is an individual of some unreal possible world. There is no object that is solely in the real world and no other, for our definition of possible worlds permits worlds that include all the objects of the real world and more.

5.4. Kripke's quantified modal logic

Saul Kripke's semantics for quantified modal logic [30] is undoubtedly the best-known formalization of possible worlds. Kripke defines a quantified model structure as a set $\mathcal{W}$ of possible worlds, one of which is the real world R, and a reflexive accessibility relation $\triangleright$ defined over the members of $\mathcal{W}$. The set $\mathcal{U}$ is the universe of individuals that turn up in at least one world, and each world W gets some subset of elements of $\mathcal{U}$ as its domain of individuals, $D(W)$.

The truth of a formula in this system is always relative to a particular world. In each world W, the *extension* of an n-ary predicate P^n, written $ext(P^n, W)$, is the set of n-tuples of individuals of $\mathcal{U}$ for which P^n is true in that world. For example, the extension of *loves* in W might be the set

$$\{\langle John, Ross\rangle, \langle Ross, Ironsides\rangle, \ldots\}.$$

Then $loves(x, y)$ would be true in W iff the pair $\langle a, b\rangle$ is in this set, where a is the individual of $\mathcal{U}$ that is assigned to the variable x in W and b is that assigned to y. Given this definition of truth for atomic formulas, operators for conjunction, $\wedge$, and negation, $\neg$, are defined in the usual way; necessity, $\Box$, is defined as truth in all accessible worlds.[26]

Note that there is no requirement that the individuals assigned to variables in W or the individuals used in the extensions of predicates in W be restricted to individuals in $D(W)$; rather, any element of $\mathcal{U}$ is allowed. Such a restriction does apply, however, in the definition of the quantifier $\forall$, which scopes only over individuals of W. That is, the formula $\forall x P^n(y_1, \ldots, y_{i-1}, x, y_{i+1}, \ldots, y_n)$ is true in W for some assignment of elements of $\mathcal{U}$ to the y_i iff it's true in W for any assignment to x from $D(W)$. So the truth in W of a sentence such as *Everyone loves Ross* is not blocked by the mere possibility that someone doesn't.

Because variables and predicate extensions can use individuals from any world, we can express propositions that relate individuals from different worlds. For example, we can say that it's true in the real world that Ross loves Pegasus, even if Ross is in the real world and Pegasus isn't. However, it's clear that Kripke himself considers this to be an infelicity of his approach. He regards it as a mere convention that such sentences have any truth value at all [30, pp. 65–66], and one might just as easily have taken the Strawsonian view (as in Woodruff's UE [77]; see above) that their value is undefined.[27] Moreover, if such sentences *are* to be given a truth value,

[26] I omit the technical details of these and other aspects of the semantics that will not concern us in this paper. The interested reader can find them in Kripke's paper [30] or textbooks such as [24, pp. 178ff].

[27] In fact, the free logic "existence" predicate, i.e., the predicate true just of terms that denote, is just the unary predicate whose extension is $D(W)$ in each W.

says Kripke [30, footnote 11], then they should always be given the value *false*! It is only for certain technical reasons (related to other concerns of Kripke's) that he does not include in his semantics the stipulation that extensions of predicates in each world W be limited to tuples of individuals in $D(W)$ [30, footnote 11].

If we take this seriously, then we are stuck. We can say that our averted strike exists as a real strike in other possible worlds, and that it has certain properties in those worlds, such as lasting for three days or three weeks. But we still can't speak truly of its properties in the worlds in which it was averted, such as its property of having been proposed by Ross, the union steward, and having been averted by the intervention of Malcolm, the mediator. But despair is unnecessary. Even if he didn't want to, Kripke has given us a formalism in which we can speak truly in one world of objects in another. We can use the formalism for what it's worth, and hope that eventually Kripke will agree that what he thought of as a bug is actually a feature.

So let's try some of our problem sentences in Kripke's logic. First, we shall assume that, in addition to worlds possible relative to the present-moment real world R, we also have worlds of the past available to us. Hence $\mathcal{U}$ includes all past objects, and we can talk about Alan Turing:

(68) Alan Turing is dead.

dead(*AlanTuring*), where the person Alan Turing is the value of the variable *AlanTuring*.

If Turing is included in *ext*(*dead*, R), then this is a well-formed sentence, true in the real world, even if Turing $\notin D(R)$. A similar treatment will work for *Alan Turing is a celebrated mathematician* and *Nadia admires Alan Turing*. And although we might feel a little worried about so doing, if we also allow Alan Turing to be in *ext*(*smart*, R), the set of objects that are smart in the present real world, then we also have *Alan Turing was smart.*

Can we also say that works of fiction are possible worlds, and thus account for dragons and Sherlock Holmes exactly as we accounted for Alan Turing? Kripke, unfortunately, objects to so doing.[28] On Kripke's view, fictional objects don't occur in any possible world; even if a world happened to contain an individual whose properties were exactly those of Sherlock Holmes, that individual would not *be* Sherlock Holmes. Alvin Plantinga [53, pp. 155–159] has also argued against a possible-world treatment of fiction. For while a possible world is complete, a fictional world is necessarily partial. For example, while it is true that Hamlet had feet, it is neither true

[28] In [30] he says otherwise, but later, in the addendum to its reprinting in [39], he explicitly repudiates this; see also [31, pp. 157–158].

nor false that his shoe size was 9B [53, p. 158]. A fiction, therefore, at best specifies a *class* of possible worlds.

But perhaps we can again ignore Kripke's advice, which is really based on a metaphysical argument as to what worlds *ought* to be considered accessible from the real world, and simply stipulate that fictional worlds *will* be considered accessible from R, and their objects *will* be in $\mathcal{U}$. The partial nature of such worlds need not concern us; it is straightforward to develop the idea of specially designated worlds in which formulas will have no truth value if they are not explicitly in accordance with, or contradicted by, the "specifications" of the world. So then we have Sherlock Holmes in $\mathcal{U}$, and it will be true, in the Sherlock Holmes world, that Sherlock Holmes was smart. It will also be true in R if we allow Sherlock Holmes to be in $ext(smart, R)$. This may come down to a matter of ontological taste.

Now let's try the cancelled lecture. Writing $\exists x$ for $\neg\forall x\neg$, we are tempted by the following:

(69) Today's lecture is cancelled.

$$\exists x \exists y\, (person(x) \wedge today's\text{-}lecture(y) \wedge cancel(x, y)).$$

But this is not correct! We want x and y to be in different worlds, but the semantics of $\exists$, for reasons crucial to the logic, requires them to both be in the world R of which we are speaking! The formula in (69) says

"There is something $x \in D(R)$, $x \in ext(person, R)$,
and something $y \in D(R)$, $y \in ext(today's\text{-}lecture, R)$,
and $\langle x, y\rangle \in ext(cancel, R)$."

While Kripke's logic allows us to *talk about* entities in other worlds, it doesn't allow us to *quantify* over them. That means that we can't pick them out by means of quantifiers and properties. For Kripke, the only way to pick out an object in another world is to use its name as a rigid designator, as we did with Alan Turing in (68). Now, Kripke does allow [31, pp. 79–80] that a suitably precise description could be a rigid designator, and this might be the case for *today's lecture* in (69), but we can't rely on this always being so:

(70) Ross cancelled one of his lectures (but I don't know which one).

So while we have, in other worlds, the lectures and strikes that didn't occur in the real world, we can't quantify over them; we can talk about them only if we have rigid designators for them—or if we reify them in the real world.

Not only can't we quantify over individuals in other worlds, but we can't quantify over the worlds themselves either, nor even refer to them explicitly, even though (as our present discussion serves to show!) they too may be objects of discourse:

(71) There are many possible worlds in which Ross is a Justice of the High Court.

(72) I have a dream of a better world, in which they are free who here are oppressed, and they are well who here are sick and lame.

The best we can do for (71) is (73), which fails to capture the meaning of *many*, as we can't talk about (the cardinality of) the set of worlds in which a proposition is true. We write $\Diamond$ ("possibly") for $\neg\Box\neg$:

(73) $\Diamond \mathit{High\text{-}Court\text{-}Justice}(\mathit{Ross})$.

"It is possible that Ross is a Justice of the High Court; there is at least one accessible world in which Ross is a Justice of the High Court."

For (72), we can use the possibility operator $\Diamond$ to implicitly invoke the possible world that is mentioned, but we again run into the problem that quantifiers scope only in a single world. The following is *not* what we want:

(74) $\Diamond\forall x((\mathit{oppressed}(x) \rightarrow \mathit{free}(x)) \wedge (\mathit{sick}(x) \rightarrow \mathit{well}(x)))$.

"There is a world in which everything that is oppressed is simultaneously free and everything that is sick is simultaneously well."

What we want to say for (72) is that everything that is oppressed in the real world R is free in the dream world, everything that is sick and lame in R is well in the other world. The problem is that we cannot, in general, write formulas in which truth in one world depends on truth in another.

We can't fix this just by following Plantinga [53, p. 47] in admitting possible worlds as objects in the universe (even though they are nonexistent!),[29] each occurring in all the worlds from which it is accessible; that is, $W' \in D(W)$ whenever $W \triangleright W'$. This doesn't help, because $W' \in D(W)$ does *not* imply that the objects in $D(W')$ are also in $D(W)$ and hence accessible to quantifiers in W. (If that were to happen, then all worlds would include all individuals.) So we still can't write a formula in W that depends on truth in W'. The following (disregarding the second part of the conjunction) still doesn't give us what we want:

(75) $$\exists W'\forall x((x \in D(R) \wedge \mathit{oppressed}(x)) \rightarrow (\exists y(y \in D(W') \wedge \mathit{free}(y) \wedge y = x))).$$

[29]Though we *model* them with mathematical objects, possible worlds are not themselves mathematical objects any more than Sherlock Holmes or the cancelled lecture are.

What we would need to carry all this through is a completely different formalization of possible worlds that would allow us to embed quantification in one world within quantification in another, indexing variables and predicates by world.

5.5. Why possible-world theories don't help

In summary, then, it seems that while the notion of possible worlds and quantified modal logics such as Kripke's might be useful mechanisms for explicating concepts of possibility and necessity, they aren't really very good with nonexistent objects. It should now be clear why this is so. The intent of Kripke's logic was to divide the universe up into separate worlds in order to constrain quantification in modal contexts, rather than to explicate the notion of nonexistent objects *per se* or to account for true assertions about objects in other worlds. (As we saw, Kripke believed that there are no such assertions.) But we weren't able to make very good use of the logic. Firstly, we used possible worlds as convenient places to store our nonexistent objects, the junk from our metaphysical attic, and not for modal reasoning at all. Secondly, we found ourselves wishing that everything would be in every world anyway, defeating the very purpose of the logic.

To put it another way, what we want to talk about and represent is one particular world, usually the actual world, and the question is therefore how dragons and averted strikes exist in the particular world of interest. It is insufficient to say merely that dragons exist in some different possible world, for so, after all, does Margaret Thatcher. That tells us nothing about the difference between dragons and Margaret Thatcher in the world that we *are* representing.

Perhaps, then, we should take courage and say that, yes, we *will* let everything be in every world and be within the scope of quantification there. Then for most purposes, we'll only need one world; it'll have everything in it that we want. Modal reasoning will still require other possible worlds—worlds in which the same universe of individuals have different properties—but that will be an orthogonal issue.

The story so far. We want to represent natural language sentences about existence and nonexistence. But construing nonexistence as existence in another possible world gets us into trouble with quantification scope and mixtures of truth in different worlds. No matter what we do, everything seems to want to collapse into one world.

Next. A solution in which everything is in one world.

6. Theories of nonexistent objects

Hobbs' scheme implicitly countenanced nonexistent objects, but, as we saw, found itself limited because it tried not to make anything special of the notion of existence. Free logics also accept nonexistent objects, but try their best to ignore them. Quantified modal logics just send them to Siberia. We now turn to an approach that doesn't just accept nonexistent objects—it whole-heartedly embraces them. The approach is that of Parsons [50]; it is explicitly motivated by Meinong's ideas (see Section 2.3 above). Parsons' goal is to define an abundant Meinongian universe that includes nonexistent objects, while excluding incoherent objects (such as those that are not self-identical) that give rise to problems and inconsistencies. [30]

Parsons defines *nuclear properties* as the "ordinary properties" that we regularly attribute to individuals [50, p. 24]. For example, being in New Zealand, being Nadia, and being Sherlock Holmes are nuclear properties, but, as we shall see, existing and being perfect are not. Corresponding to each nuclear property is a *nuclear predicate* that is true of the individuals that have that property. There are also *nuclear relations* of two (or more) places; for example, Nadia and her cat may be in the nuclear relationship that the former feeds the latter.

In Parsons' theory, for each distinct set of nuclear properties, the unique object that has exactly that set of properties is included in the universe over which quantifiers scope. But that's all that's in the universe. There is an object that is green (and has no other nuclear property but that); there is an object that is both green and Nadia; there is even an object that is green and Nadia and Sherlock Holmes. But not all these objects exist in the real world—in some cases because they just happen not to, and in other cases because they are not possible.

Properties and relations that aren't nuclear are said to be *extranuclear*. The prime example is physical existence, written *E!*. Thus, existence is taken as a predicate, but one of a special kind. Some other extranuclear predicates are: being perfect, being possible, being an object in the universe, being worshipped by Ross, and being thought about by Margaret Thatcher. (However, worshipping Zeus and thinking about Margaret Thatcher would

[30] Rapaport [59,60,62] has also presented a Meinong-inspired theory of nonexistent objects. Space does not permit discussion of both theories. The main differences between the two are the following:

(1) Parsons has only one type of object, which may or may not exist, whereas Rapaport distinguishes Meinongian objects ("M-objects") from actual objects ("*sein*-correlates" of M-objects).

(2) Parsons has two types of predicate, whereas Rapaport has one type that can be applied in two different ways: actual objects "exemplify" their properties, whereas M-objects "are constituted" by their properties.

both be nuclear.) Parsons admits [50, p. 24] to being unable to precisely characterize the distinction between the two types of predicate. He suggests, however, that in any particular case, if there is any doubt or controversy over whether a particular property or relation is nuclear, then it probably isn't. Another clue comes from the fact that nuclear relations may hold only between two existent objects or between two nonexistent objects; any relation that can hold between an existent and a nonexistent object must be extranuclear [50, p. 160]. Thus, *is-taller-than* is an extranuclear relation, because Margaret Thatcher (who exists) is taller than Hercule Poirot (who doesn't) [50, pp. 168–169]. In fact, by a similar argument, any comparative relation is extranuclear, and so are relations like *avert* and *cancel.* [31]

Although the universe is defined in terms of distinct sets of nuclear properties, any object in the universe may also have extranuclear properties. In fact, they all have the extranuclear property of being an object in the universe, for example; and some have the extranuclear property of physical existence.

Now, the tricky part is what to do with objects like the golden mountain and the existent golden mountain. These both have exactly the same set of nuclear properties, i.e., goldenness and mountainhood, and are therefore the "same" object by our earlier definition. This seems undesirable; intuitively, "the X" and "the existent X" are different objects—especially if X isn't itself existent. Yet the existent golden mountain must be accounted for, as we can still talk about it, and the account must not entail its existence. So following Meinong, Parsons introduces the concept of *watering down* extranuclear properties to nuclear ones. Thus for Parsons, there is also an existence property that's nuclear—call it $E!_N$. That's the kind of existence that the existent golden mountain has, and that's how it gets into the universe as a distinct object from the regular golden mountain. Watered-down existence says nothing about real, genuine, full-blown extranuclear existence, and the existent golden mountain still doesn't have the latter. A similar story can be told about the possible round square; its possibility is merely the watered-down variety.

The watering-down operation on an extranuclear predicate creates a new nuclear predicate that among existing objects is true of the same objects of which the original predicate was true. That is, if a given existing object has an extranuclear predicate true of it, it will have the corresponding watered-down nuclear predicate true of it as well; and vice versa. Anything

[31] In his formalization, to be discussed below, Parsons excludes extranuclear relations, such as *worship*, *avert*, and *cancel*, that yield a nuclear property when one of their argument positions is closed ("plugged up") and an extranuclear property when the other one is. He claims [50, p. 65] that this is for simplicity, and that there are no theoretical difficulties in including such relations. In Section 7.2, we shall rely on this indeed being so, and assume them to have been added to the formalization.

that exists full-strength also exists in a watered-down way; anything that exists that is full-strength-possible is also watered-down-possible. Among nonexistent objects, however, the extranuclear predicate and its watered-down counterpart may diverge. But it's not clear just what sort of a thing these watered-down properties are. What exactly is it that the watered-down-existent gold mountain has that the regular gold mountain doesn't? Just, it seems, an abstract attribution that has no effect on anything except in serving to distinguish the two.

Parsons develops a formal language, called $\mathcal{O}$, for talking about this universe. $\mathcal{O}$ is a second-order modal language with belief contexts; quantification is explicitly over all objects in the universe. The language distinguishes the two types of predicates, and the extranuclear predicate of existence, $E!$, has special axiomatic properties. The watering-down operation on extranuclear predicates is defined. The modalities of necessity and possibility are defined over a set of possible universes; but each possible universe contains the same objects and differs from the others only in which objects have which properties (including existence). Using Montague-like techniques [46], Parsons shows how $\mathcal{O}$ can act as a semantics for a fragment of English, treating sentences such as:

(76) The King of France doesn't exist.

$$\neg(\iota x)(E!(x) \wedge \mathit{King\text{-}of\text{-}France}(x))[\lambda y E!(y)].$$

Roughly, this says that it is not true that there is—in the actual world—a unique x that both is the King of France and exists in the world; if there is indeed no King of France, this formula is true. Also included in the fragment is the sentence *Every good modern chemist knows more about chemical analysis than Sherlock Holmes* (cf. sentence (v) of footnote 10).

If we are willing to accept Parsons' approach, then a number of our problems are solved. We can talk about Sherlock Holmes and dragons and other fictional objects all we like. (Parsons devotes two chapters to fictional objects.) We also have Alan Turing available, and, presumably, all future objects. And we have lots of useful objects that don't exist, including strikes and lectures that never happened—that is, we have the objects that have exactly the properties required, with no necessity that they exist. And the existence of God is not a theorem, no matter how God is described; "for either the description will be purely nuclear in character, and we will not be able to show that the objects [that] satisfy it exist, or it will be partially extranuclear, and we will not be able to show that *any* object [in the universe] satisfies it" ([50, p. 213], emphasis added).

It should be noted, however, that by the same argument, we are not actually *guaranteed* to have averted strikes or cancelled lectures *per se* in the universe, because being averted and being cancelled are extranuclear properties. What we do have at least are strikes and lectures that have all the exact same nuclear properties as the strikes and lectures of interest, including strikes that have been watered-down-averted and lectures that have been watered-down-cancelled. Whether any particular strike or lecture is genuinely, extranuclearly averted or cancelled will be a matter of contingent fact.

Parsons' approach is not without problems. (See Rapaport [64] for a detailed critique.) For example, while nonexistent strikes and lectures are available as objects, we can't do everything with them that we would like. $\mathcal{O}$ can say that an existent Ross stands in a *cancelled* relation to a nonexistent lecture, but it is not possible, I think, to explicate the meaning of this as Ross *causing* the nonexistence; Parsons did not consider such things.

Another problem is the profligate scope of the quantifiers. An insight from free logic and Kripke's quantified modal logic that must be retained is that quantification scope must be restrained. Parsons' universe is much too large to quantify over, because it contains a counterexample to every nuclear proposition, an instance of every set of nuclear properties. For example, in Parsons' universe, the sentences *No pigs fly* and *All marmots are mortal* are false, because the universe includes flying pigs and immortal marmots. The effect is rather like that of the Sorcerer's Apprentice; we wanted to account for just a few nonexistent objects, and now we find hordes of them coming out of the woodwork like cockroaches.

But there is no single correct constraint on quantification. For example, it would normally be silly to quantify over all the unwritten books, unthought ideas, or unlived lives; but sometimes, one might have to do so. (An unwritten book is surely reified in the sentence *Ross is going to start writing a book.*) In KR systems, this may not be a practical problem, for the size of the universe is limited by the size of the knowledge base anyway, and even within that, searches would normally be further constrained. This is not to say that a knowledge base cannot contain (finite representations of) infinite objects—the set of integers, for example—but a practical system will normally limit itself to the entities it already knows about and won't capriciously start generating new ones just to see what turns up.

Despite these problems, we'll see in Section 7.2 below that a number of aspects of Parsons' approach are helpful in our goal of including nonexistent objects in a knowledge representation formalism.

The story so far. We want to represent natural language sentences about existence and nonexistence. But knowledge representation formalisms either impute existence to objects when they shouldn't, or they get into trouble treating existence as a predicate. Free logics and possible-world theories don't help either. Philosophical theories of nonexistent objects offer some hope for a solution.

Next. Naivety to the rescue.

7. Naive ontology: the ontology of natural language

Let's take stock of where we are. We've seen three separate ideas of what the set of things that exist is:

A: the things that physically exist (plus mathematical objects);
B: the things that quantifiers scope over;
C: the things we can talk and think about.

We've seen these ideas related in various ways. The austere view, from Russell and Quine, is that $A = B = C$. The promiscuous view, from Meinong and Hobbs, is that $A \subset B = C$. In between, Kripke and Parsons, in different ways, say that $A \subset B \subset C$—that is, they try to be as promiscuous as possible without actually getting into trouble. I've argued throughout the paper that a generally promiscuous approach is required for an adequate representation of natural language in AI. In this section, now, I want to lay the foundation for such a representation. I'll be taking the promiscuous-but-cautious view, $A \subset B \subset C$, making B as large as possible.

7.1. *Different kinds of existence*

The real problem with the Russell–Quine position, the free-logic and possible-world approaches, and even Parsons' approach is that they equivocate about existence; they speak as if all things that exist exist in the same way. This is clearly not so. Margaret Thatcher exists, and so does the number 27, but they do so in different ways: one is a physical object in the world, while the other has only abstract existence. But even Quine is willing to grant the existence of mathematical entities—and of concepts in general. If we admit these two kinds of existence, then perhaps we can find even more kinds if we look. And arguments about the nature of one kind—whether it can be a predicate, for example—need not hold true of the others.

In fact, following the style of naive physics [17], we can develop a *naive ontology* that captures the commonsense view of existence that natural language reflects. In doing so, we follow Meinong in not limiting membership in the universe to things in the world, but attributing it to anything that can be spoken of. The commonsense notion that anything that can be spoken of has being of some kind or another may not stand up to intense scrutiny, but is certainly robust enough for our naive approach. (Plantinga [53], for example, shows that the notion is able to withstand quite a number of philosophical challenges, and needs to go to some length before he believes that he can claim that he has defeated it.)

And we go further, by imposing a taxonomy of existence upon the universe, identifying about eight different kinds of existence. In particular, we solve the problems of the cancelled lecture and the averted strike by attributing some kind of being to them (but not physical actuality). Thus all sentences will be about objects that are somewhere in the universe, and will therefore have the potential to be true.

We start by taking the universe to be as Parsons defined it: the set of objects given by all possible distinct nonempty combinations of nuclear properties, including watered-down extranuclear properties. This will give us a large assortment of physical objects, mathematical objects, concepts, and so on. This is the kosher part of the universe. To this, we add a "quarantine" section in which objects live with no nuclear properties at all. These are the *tref* objects that would create inconsistency in Parsons' system: Russell sets, non-self-identical objects, and so on. The various kinds of existence that we identify, all in the kosher part of this universe, are then as follows. All are extranuclear properties:

- Physical existence in the present real world (or that under consideration), with causal interaction. Margaret Thatcher exists this way, and so do events such as Nadia's putting the cat out. This is the same property as that of Parsons' original *E*! predicate.
- Physical existence in a past world (with causal interaction therein, and some indirect causal connection to the present world). The late Alan Turing, for example, exists in a world of the past; he doesn't exist now, but nevertheless he *is*, in the present, a celebrated mathematician, and likewise he *is* dead (see Section 3.2.2 above).
- Abstract, necessary existence, as of mathematical objects such as 27 and the least prime greater than 27.
- Existence outside a world, but with causal interaction with that world. This is the kind of existence that most Western religions attribute to God.
- Abstract, contingent existence in the real world. Freedom, suavity, and fear would come into this category.
- Existence as a concept, which is abstract but contingent, such as the concept of Margaret Thatcher, which need not have existed. [32]

[32]One may wish to combine this category with the previous one, saying that concepts are not ontologically distinct from other abstract entities like suavity. I will not take a position on this. Alternatively, one might argue that the existence of a concept may be necessary or contingent depending on its extension. That is, the concept of Margaret Thatcher is as contingent as Margaret Thatcher is, but the concept of the least prime greater than 27 is necessary because its extension is. The category of existence as a concept would then be split over the three abstract categories above.

- Unactualized existence.[33] This category includes objects that could become actual in the future, objects in counterfactuals, "past" objects that never came into being, and perhaps also impossible objects. Strictly speaking, this category crosses with the previous six. The baby that Diane wants to have has unactualized physical existence; the book that Ross once wanted to write has unactualized past existence; and hypothetical gods have unactualized divine existence. It's not clear to me that unactualized necessary existence is meaningful, unless that's the kind that $\sqrt{-1}$ has. Note that objects in the quarantine section of the universe do not have even unactualized existence.
- Existence in fiction. This is the sense in which Sherlock Holmes and dragons exist.[34] This category, too, crosses with the others. Sherlock Holmes and dragons have fictional physical existence; mythological gods have fictional divine existence; and a story about a counterexample to the four-color theorem invokes fictional necessary existence.[35]

My point here is not to argue for exactly this list of types of existence—that's a topic in philosophy, not artificial intelligence—but rather to demonstrate that however many distinct types of existence there are, it's somewhat more than two.[36] Any knowledge representation formalism that is to be adequate to the task of natural language understanding will need to be able to account for them all—that is, it will treat existence as a set of properties, and, given a particular object's mode of existence, draw inferences accordingly.

It should be clear that the various kinds of existence can't all be accounted for just by organizing the IS-A hierarchy the right way. It is true that one can, at the top, make a distinction between abstract and concrete entities. But

[33]I use this horrible term for want of a better one.

[34]"Everyone knows that dragons don't exist. But while this simplistic formulation may satisfy the layman, it does not suffice for the scientific mind. ... The brilliant Cerebron, attacking the problem analytically, discovered three distinct kinds of dragon: the mythical, the chimerical, and the purely hypothetical. They were all, one might say, nonexistent, but each nonexisted in an entirely different way." (Stanisław Lem [33, p. 76])

[35]This still leaves a few loose ends. For example, it could be argued that the fictional physical existence of, say, Sherlock Holmes entails both the fictional existence and the actual existence of the concept of Sherlock Holmes. Are these then two separate entities, or one entity with a dual mode of existence, or what?

[36]Routley [65, p. 441] objects to all "kinds-of-existence doctrines", apparently because they don't have the guts to come right out and say that there are things that just plain don't exist. Routley puts his position by parody rather than argument ("canned peaches exist as grocery supplies"), so his objections remain unclear. But it seems to me that if there is a dispute, it is terminological; to the optimist, an object has "unactualized existence", while to the pessimist, it's simply "nonexistent". Moreover, I think Routley's objections are misdirected. His main aim is to attack the "ontological assumption"—basically, a bias against nonexistence. But our naive ontology here does not include the dreaded ontological assumption, and indeed is consistent with its converse (cf. [61, p. 550n]).

past existence, unactualized existence, and fictional existence are certainly orthogonal to the hierarchy of concrete entities. And it is usual to arrange an IS-A hierarchy as a network in which nodes representing instances are necessarily leaves and those representing concepts are (or can be) interior nodes; there are clear advantages in retaining this structure for reasoning about inheritance of properties, rather than trying to separate concepts and instances as fundamentally different types.

7.2. Using the naive ontology

We can now show how the naive ontology can be used to fix some of the problems of transparency and entailment of existence in Hobbs' system. I will not present a formalization, as many details remain to be worked out.

First, recall that in Parsons' system, nuclear relations could hold only between objects that both existed or both didn't. We can immediately generalize this: nuclear relations may hold only between objects that exist in *the same way*. For example, *instance-of* will not be nuclear, as it can relate concepts, which exist one way, to objects that exist in other ways. As before, *avert*, *cancel*, and so on will also be extranuclear.

Second, we take the notion of watering-down to mean severely weakening a predicate to the point where it becomes nothing but an abstract attribute with no significant consequences. We do this by prohibiting watered-down properties from entailing anything but other watered-down properties. So, for example, while the extranuclear property of omniscience entails the nuclear property of knowing where Ross is, watered-down omniscience does not.

Third, we prohibit objects in the quarantined section of the universe from doing just about everything. Intuitively, we allow them to be mentioned, but not used. So we can talk about Russell sets, and our use of the term will refer, but that's about all. They may not appear in any axiom, nor participate in any inference. And quantifiers do not scope over them.[37] (Note that these restrictions do not apply to the *concepts* of the *tref* objects; these have healthy, conceptual existence in the kosher section of the universe.)

Next, we extend the notion of transparent argument positions, as in Hobbs' system, so that the existence of various objects can be inferred from assertions about relationships in which they participate. Let's consider simple nuclear relationships first:

(77) Ross kisses Nadia.

kiss(*Ross*, *Nadia*).

[37]Thus with respect to quantifiers, these objects are rather like objects in other possible worlds in Kripke's quantified modal logic (Section 5.4 above); that is, they can be picked out by a rigid designator but not by a quantifier.

Because it is nuclear, both argument positions of *kiss* will be transparent. From this, we will now infer *not* that Nadia and Ross exist, but rather that to the extent that they exist, they do so in the same way—both are physical or fictional or past or whatever. (Presumably real-world physical existence would be a good default assumption if there were no indication to the contrary; and the various kinds of conceptual, abstract, and divine existence would be ruled out by the lexical semantics of *kiss*.) Using Hobbs' style of formalism, we can go further. We must revise Hobbs' basic form, which was (78):

(78) $kiss'(E, Ross, Nadia) \wedge Exist(E)$

to specify what kind of existence E has. It will then follow that Ross and Nadia exist the same way that E does; for example, in (79):

(79) $kiss'(E, Ross, Nadia) \wedge Physically\text{-}exist(E)$

we can infer the physical existence of Ross and Nadia from that of the kissing action.

In the case of extranuclear relationships, such inferences do not go through. As desired, we can infer nothing from (80) about the ontological status of Margaret or Hercule:

(80) $taller\text{-}than(Margaret, Hercule)$.

But some extranuclear relationships admit what we earlier (Section 4.1) called *anti-transparent positions*. Our paradigm case is the averted strike. Even if averting *per se* is extranuclear, the property of being an act of averting seems to be nuclear:

(81) $avert'(E, Ross, Strike) \wedge Exist(E)$,

where *Exist* is now taken to mean existing in one way or another. Because *avert* is extranuclear, no inferences can be automatic here. Rather, it is a matter of the lexical semantics of *avert* that certain limited inferences go through: that Ross exists the way E does and that the strike must have unactualized existence.

Lastly, we are protected against accidentally defining God, or anything else, into real existence. The assertion of a nuclear property allows one to infer only that the individual of which it is predicated exists in some way. For example, the truth of $green(Nadia)$ doesn't entail Nadia's physical existence, but only that she is in the universe somewhere. And the assertion of an extranuclear property does still less; the truth of $perfect(God)$ doesn't entail that God is even in the unquarantined universe. (A watered-down-perfect God is, but nothing interesting follows from that.)

8. Conclusion

What I've shown in this paper is that knowledge representation formalisms that are to be suitable for use in natural language understanding must take account of the ways that existence and nonexistence can be spoken of in natural language. Neither the traditional approaches of Frege, Russell, and Quine, nor possible-world theories and free logics are adequate.

Intuitively, a better approach seems to require treating existence as a predicate and including nonexistent objects in the universe over which our quantifiers scope—much as Hobbs did. Philosophers have traditionally taken a dim view of such activities, however, and I've tried to show the reasons for their concern. Nevertheless, I think Hobbs' approach is the most promising of those that we've looked at. But developing it further requires developing the notion of a naive ontology. The task is analogous to naive physics and other projects in AI to represent commonsense notions of the world, and in this paper, I've presented a first cut at such an ontology and shown how it could be added to Hobbs' system.

I also see promise in Parsons' Meinongian account. By basing our definition of the universe on his, we were able to give our naive ontology a large supply of useful objects without it lapsing into inconsistency. And Parsons' distinction between nuclear and extranuclear predicates can help strengthen a Hobbs-like approach against the wrath of the philosophers that it scorns. In Section 7.2, I've sketched an outline of how the distinction could be used.

There are many details left to be worked out, of course. However, I will have succeeded in my goals for this paper if I have convinced the reader that nonexistent objects, their representation, and their role in quantification are important concerns in artificial intelligence, but there are no workable, off-the-shelf solutions in philosophy that we can just take and use.

Acknowledgement

A small, early, less-correct version of this paper was presented at the First International Conference on Principles of Knowledge Representation and Reasoning, Toronto, May 1989, and appears in the Proceedings thereof [5].

For discussions on these matters and comments and advice on earlier versions of this paper, I am grateful to Stephen Regoczei, Bill Rapaport, Jeff Pelletier, James Allen, John Barnden, Stephen Bellantoni, Ed Cohen, Chrysanne DiMarco, Jerry Hobbs, Diane Horton, Shalom Lappin, Andrew Malton, Chris Mellish, Terence Parsons, Anonymous Referees, Allen Renear, Barry Richards, Graeme Ritchie, Stuart Shapiro, John Sowa, Joan Talent, Nadia Talent, Eric Wheeler, Bill Woods, Felix Yen, and Włodek Zadrożny.

The foundations of this work were laid while I was at the Department of

Artificial Intelligence, University of Edinburgh, with the support of a Visiting Fellowship from the U.K. Science and Engineering Research Council. The balance of the work was supported by the Natural Sciences and Engineering Research Council of Canada. The final revision was prepared with facilities kindly provided by the Department of Computer Science, University of Rochester, while I was a visitor there.

References

[1] Anselm, *Proslogion* (M.J. Charlesworth, translator) (Clarendon Press, Oxford, 1965).

[2] J. Barnes, *The Ontological Argument* (Macmillan, London, 1972).

[3] R.J. Brachman, R.E. Fikes and H.J. Levesque, KRYPTON: a functional approach to knowledge representation, Tech. Report 16, Fairchild Laboratory for Artificial Intelligence Research, Palo Alto, CA (1983); reprinted in: R.J. Brachman and H.J. Levesque, eds., *Readings in Knowledge Representation* (Morgan Kaufmann, Los Altos, CA, 1985) 411–429.

[4] R.J. Brachman and H.J. Levesque, eds., *Readings in Knowledge Representation* (Morgan Kaufmann, Los Altos, CA, 1985).

[5] R.J. Brachman, H.J. Levesque and R. Reiter, eds., *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Ont. (Morgan Kaufmann, San Mateo, CA, 1989).

[6] E. Charniak, M.K. Gavin and J.A. Hendler, The Frail/NASL reference manual, Tech. Report CS-83-06, Department of Computer Science, Brown University, Providence, RI (1983).

[7] R.M. Chisholm, Identity through possible worlds: some questions, *Noûs* **1** (1968) 1–8.

[8] G. Currie, *Frege: An Introduction to His Philosophy*, Harvester Studies in Philosophy **11** (Harvester, Brighton, 1982).

[9] D. Davidson, *Essays on Actions and Events* (Clarendon Press, Oxford, 1980).

[10] R. Descartes, Meditations on first philosophy, in: *The Philosophical Works of Descartes* Vol. I (E.S. Haldane and G.R.T. Ross, translators) (Cambridge University Press, Cambridge, 1911).

[11] G. Fauconnier, *Mental Spaces: Aspects of Meaning Construction in Natural Language* (MIT Press, Cambridge, MA, 1985).

[12] H. Feigl and W. Sellars, eds., *Readings in Philosophical Analysis* (Appleton-Century-Croft, New York, 1949).

[13] G. Frege, *Die Grundlagen der Arithmetik* (Wilhelm Köbner, Breslau, 1884); reprinted as: G. Frege, *The Foundations of Arithmetic* (J.L. Austin, translator) (Blackwell, Oxford, 2nd rev. ed., 1953).

[14] G. Frege, Über Sinn und Bedeutung, *Z. Philos. Philos. Kritik* **100** (1892) 25–50; reprinted as: On sense and nominatum (H. Feigl, translator), in: H. Feigl and W. Sellars, eds., *Readings in Philosophical Analysis* (Appleton-Century-Croft, New York, 1949) 85–102; also as: On sense and reference (M. Black, translator), in: P.T. Geach and M. Black, eds., *Translations from the Philosophical Writings of Gottlob Frege* (Blackwell, Oxford, 1952) 56–78.

[15] G. Frege, *On the Foundations of Geometry and Formal Theories of Arithmetic* (E.-H.W. Kluge, translator) (Yale University Press, New Haven, CT, 1971).

[16] P.T. Geach and M. Black, eds., *Translations from the Philosophical Writings of Gottlob Frege* (Blackwell, Oxford, 1952).

[17] P.J. Hayes, The second naive physics manifesto, in: J.R. Hobbs and R.C. Moore, eds., *Formal Theories of the Commonsense World* (Ablex, Norwood, NJ, 1985) 1–36; reprinted in: R.J. Brachman and H.J. Levesque, eds., *Readings in Knowledge Representation* (Morgan Kaufmann, Los Altos, CA, 1985) 468–485.

[18] G. Hirst, *Semantic Interpretation and the Resolution of Ambiguity* (Cambridge University Press, Cambridge, 1987).
[19] G. Hirst, Semantic interpretation and ambiguity, *Artif. Intell.* **34** (1988) 131–177.
[20] G. Hirst, Knowledge representation problems for natural language understanding, Abstract in: H. Trost, ed., *4. Österreichische Artificial-Intelligence-Tagung: Wiener Workshop Wissensbasierte Sprachverarbeitung: Proceedings*, Informatik-Fachberichte **176** (Springer, Berlin, 1988).
[21] J.R. Hobbs, Ontological promiscuity, in: *Proceedings 23rd Annual Meeting of the Association for Computational Linguistics*, Chicago, IL (1985) 61–69.
[22] D.L. Horton, Incorporating agents' beliefs in a model of presupposition, M.Sc. Thesis, Tech. Report CSRI-201, Computer Systems Research Institute, University of Toronto, Toronto, Ont. (1987).
[23] D.L. Horton and G. Hirst, Presuppositions as beliefs, in: *Proceedings International Conference on Computational Linguistics (COLING-88)*, Budapest (1988) 255–260.
[24] G.E. Hughes and M.J. Cresswell, *An Introduction to Modal Logic* (Methuen, London, 1968).
[25] D. Hume, *Dialogues Concerning Natural Religion* (N. Kemp Smith, ed.) (Clarendon Press, Oxford, 1935).
[26] I. Kant, *Critique of Pure Reason* (N. Kemp Smith, translator) (Macmillan, London, 2nd rev. ed., 1933).
[27] W.E. Kennick. Moore on existence and predication, in: A. Ambrose and M. Lazerowitz, eds., *G.E. Moore: Essays in Retrospect* (Allen & Unwin, London, 1970) 160–192.
[28] M. Koubarakis, J. Mylopoulos, M. Stanley and M. Jarke, TELOS: a knowledge representation language for requirements modelling, Tech. Report KRR-89-1, Department of Computer Science, University of Toronto, Toronto, Ont. (1989).
[29] M. Koubarakis, J. Mylopoulos, M. Stanley and A. Borgida, TELOS: features and formalization, Tech. Report KRR-89-4, Department of Computer Science, University of Toronto, Toronto, Ont. (1989).
[30] S.A. Kripke, Semantical considerations on modal logic, *Acta Philos. Fenn.* **16** (Proceedings of a colloquium on modal and many-valued logics) (1963) 83–94; reprinted, with addendum, in: L. Linsky, ed., *Reference and Modality* (Oxford University Press, Oxford, 1971) 63–72 and 172.
[31] S.A. Kripke, *Naming and Necessity* (Harvard University Press, Cambridge, MA, 1980).
[32] K. Lambert, *Meinong and the Principle of Independence* (Cambridge University Press, Cambridge, 1983).
[33] S. Lem, The third sally, or The dragons of probability, in: *The Cyberiad: Fables for the Cybernetic Age* (M. Kandel, translator) (Avon Books, New York, 1976).
[34] S.C. Levinson. *Pragmatics* (Cambridge University Press, Cambridge, 1983).
[35] D. Lewis, Counterpart theory and quantified modal logic, *J. Philos.* **65** (1968) 113–126; reprinted, with addenda, in: D. Lewis, *Philosophical Papers*, Vol. I (Oxford University Press, New York, 1983) 26–46.
[36] D. Lewis, Anselm and actuality, *Noûs* **4** (1970) 175–188; reprinted, with addenda, in: D. Lewis, *Philosophical Papers*, Vol. I (Oxford University Press, New York, 1983) 10–25.
[37] D. Lewis, *Philosophical Papers*, Vol. I (Oxford University Press, New York, 1983).
[38] D. Lewis and S. Lewis, Holes, *Australasian J. Philos.* **48** (1970) 206–212; reprinted in: D. Lewis, *Philosophical Papers*, Vol. I (Oxford University Press, New York, 1983) 3–9.
[39] L. Linsky, ed., *Reference and Modality* (Oxford University Press, Oxford, 1971).
[40] A.S. Maida and S.C. Shapiro, Intensional concepts in propositional semantic networks, *Cognitive Sci.* **6** (1982) 291–330.
[41] J. Margolis, ed., *An Introduction to Philosophical Enquiry* (Knopf, New York, 1968).
[42] J.N. Martin, *Elements of Formal Semantics: An Introduction to Logic for Students of Language* (Academic Press, Orlando, FL, 1987).
[43] J. McCarthy, Epistemological problems of artificial intelligence, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 1038–1044.
[44] J. McCarthy, First order theories of individual concepts and propositions, in: J.E. Hayes,

D. Michie and L.I. Mikulich, eds., *Machine Intelligence* **9** (Ellis Horwood, Chichester, 1979) 129–147; reprinted in: R.J. Brachman and H.J. Levesque, eds., *Readings in Knowledge Representation* (Morgan Kaufmann, Los Altos, CA, 1985) 523–533.

[45] A. Meinong, Über Gegenstandstheorie, in: A. Meinong, ed., *Untersuchungen zur Gegenstandstheorie und Psychologie* (Barth, Leipzig, 1904); reprinted in: A. Meinong, *Gesamtausgabe*, Vol. II (R. Haller and R. Kindinger, eds.) (Akademische Druck- und Verlaganstalt, Graz, 1969–1978) 481–535; reprinted as: The theory of objects (I. Levi, D.B. Terrell and R.M. Chisholm, translators), in: R.M. Chisholm, ed., *Realism and the Background of Phenomenology* (Free Press, Glencoe, IL, 1960) 76–117.

[46] R. Montague, The proper treatment of quantification in ordinary English, in: K.J.J. Hintikka, J.M.E. Moravcsik and P.C. Suppes, eds., *Approaches to Natural Language: Proceedings 1970 Stanford Workshop on Grammar and Semantics* (Reidel, Dordrecht, 1973) 221–242; reprinted in: R. Montague, *Formal Philosophy: Selected Papers of Richard Montague* (R.H. Thomason, ed.) (Yale University Press, New Haven, CT, 1974) 247–270.

[47] G.E. Moore, Is existence a predicate? in: *What Can Philosophy Determine?*, Proceedings Aristotelian Society, Suppl. Vol. **15** (1936) 175–188; reprinted in: G.E. Moore, *Philosophical Papers* (Allen & Unwin, London, 1959), 115–126; also in: A.G.N. Flew, ed., *Logic and Language* (Second Series) (Basil Blackwell, Oxford, 1953) 82–94.

[48] S. Neale, *Descriptions* (MIT Press, Cambridge, MA, 1990).

[49] T. Parsons, Essentialism and quantified modal logic, *Philos. Rev.* **78** (1969) 35–52; reprinted in: L. Linsky, ed., *Reference and Modality* (Oxford University Press, Oxford, 1971) 73–87.

[50] T. Parsons, *Nonexistent Objects* (Yale University Press, New Haven, CT, 1980).

[51] P.F. Patel-Schneider, A four-valued semantics for frame-based description languages, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 344–348.

[52] A. Plantinga, *God and Other Minds: A Study of the Rational Justification of Belief in God* (Cornell University Press, Ithaca, NY, 1967).

[53] A. Plantinga, *The Nature of Necessity* (Clarendon Press, Oxford, 1974).

[54] A.N. Prior, Existence, in: P. Edwards, ed., *The Encyclopedia of Philosophy* **3** (Macmillan, New York, 1967) 141–147.

[55] R.L. Purtill, About identity through possible worlds, *Noûs* **2** (1968) 87–89.

[56] W.V.O. Quine, On what there is, *Rev. Metaphys.* **1** (1948) 21–38; reprinted in: *Freedom, Language, and Reality*, Proceedings Aristotelian Society, Suppl. Vol. **25** (1951) 216–233; also in: W.V.O. Quine, *From a Logical Point of View: Logico-Philosophical Essays* (Harvard University Press, Cambridge, MA, 1st ed., 1953; 2nd ed., 1961) 1–19; also in: J. Margolis, ed., *An Introduction to Philosophical Enquiry* (Knopf, New York, 1968) 668–679.

[57] W.V.O. Quine, Reference and modality, in: W.V.O. Quine, *From a Logical Point of View: Logico-Philosophical Essays* (Harvard University Press, Cambridge, MA, 1st ed., 1953; 2nd ed., 1961) 139–157; reprinted in: L. Linsky, ed., *Reference and Modality* (Oxford University Press, Oxford, 1971) 17–34.

[58] W.V.O. Quine, *From a Logical Point of View: Logico-Philosophical Essays* (Harvard University Press, Cambridge, MA, 1st ed., 1953; 2nd ed., 1961).

[59] W.J. Rapaport, Meinongian theories and a Russellian paradox, *Noûs* **12** (1978) 153–180.

[60] W.J. Rapaport, How to make the world fit our language: an essay in Meinongian semantics, *Grazer Philos. Stud.* **14** (1981) 1–21.

[61] W.J. Rapaport, Review of: R. Routley, Exploring Meinong's jungle and beyond: an investigation of noneism and the theory of items (interim ed.), Departmental Monograph 3, Philosophy Department, Research School of Social Sciences, Australian National University, Canberra, ACT (1980), *Philos. Phenomenological Res.* **44** (1984) 539–552.

[62] W.J. Rapaport, Non-existent objects and epistemological ontology, *Grazer Philos. Stud.* **25/26** (1985/86) 61–95.

[63] W.J. Rapaport, Meinongian semantics for propositional semantic networks, in: *Proceedings 23rd Annual Meeting of the Association for Computational Linguistics*,

Chicago, IL (1985) 43–48.

[64] W.J. Rapaport, To be and not to be (Review of: T. Parsons, *Nonexistent Objects* (Yale University Press, New Haven, CT, 1980)), *Noûs* **19** (1985) 255–271.

[65] R. Routley, Exploring Meinong's jungle and beyond: an investigation of noneism and the theory of items (interim ed.), Departmental Monograph 3, Philosophy Department, Research School of Social Sciences, Australian National University, Canberra, ACT (1980).

[66] B. Russell, On denoting, *Mind n.s.* **14** (1905) 479–493; reprinted in: B. Russell, *Essays in Analysis* (D. Lackey, ed.) (Allen & Unwin, London, 1973) 103–119; reprinted in: H. Feigl and W. Sellars, eds., *Readings in Philosophical Analysis* (Appleton-Century-Croft, New York, 1949) 103–115; also in: B. Russell, *Logic and Analysis: Essays 1901–1950* (R.C. Marsh, ed.) (Allen & Unwin, London, 1956) 39–56; also in: J. Margolis, ed., *An Introduction to Philosophical Enquiry* (Knopf, New York, 1968) 631–642.

[67] B. Russell, The philosophy of logical atomism, in: B. Russell, *Logic and Analysis: Essays 1901–1950* (R.C. Marsh, ed.) (Allen & Unwin, London, 1956) 175–281.

[68] B. Russell, *Logic and Analysis: Essays 1901–1950* (R.C. Marsh, ed.) (Allen & Unwin, London, 1956).

[69] G. Ryle, Systematically misleading expressions, *Proceedings Aristotelian Society* **32** (1931–32) 139–170; reprinted in: A.G.N. Flew, ed., *Logic and Language* (First Series) (Basil Blackwell, Oxford, 1950) 11–36.

[70] R. Schock, *Logics without Existence Assumptions* (Almqvist & Wiksell, Stockholm, 1968).

[71] S.C. Shapiro and W.J. Rapaport, SNePS considered as a fully intensional propositional semantic network, in: N. Cercone and G. McCalla, eds., *The Knowledge Frontier: Essays in the Representation of Knowledge* (Springer, New York, 1987) 262–315.

[72] R.M. Smullyan, *What is the Name of This Book?—The Riddle of Dracula and Other Logical Puzzles* (Prentice-Hall, Englewood Cliffs, NJ, 1978).

[73] H.D. Sluga, *Gottlob Frege* (Routledge & Kegan Paul, London, 1980).

[74] J.F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine* (Addison-Wesley, Reading, MA, 1984).

[75] P.F. Strawson, On referring, *Mind n.s.* **59** (1950) 320–344.

[76] C.J.F. Williams, *What is Existence?* (Clarendon Press, Oxford, 1981).

[77] P.W. Woodruff, Logic and truth value gaps, in: K. Lambert, ed., *Philosophical Problems in Logic: Some Recent Developments* (Reidel, Dordrecht, 1970) 121–142.

[78] J. Woods, *The Logic of Fiction: A Philosophical Sounding of Deviant Logic*, De Proprietatibus Litterarum, Series Minor **16** (Mouton, The Hague, 1974).

[79] W.A. Woods, What's in a link: foundations for semantic networks, in: D.G. Bobrow and A.M. Collins, eds., *Representation and Understanding: Studies in Cognitive Science, Language, Thought and Culture: Advances in the Study of Cognition* (Academic Press, New York, 1975) 35–82; reprinted in: R.J. Brachman and H.J. Levesque, eds., *Readings in Knowledge Representation* (Morgan Kaufmann, Los Altos, CA, 1985) 217–241.

[80] E.N. Zalta, *Abstract Objects: An Introduction to Axiomatic Metaphysics*, Synthese Library **160** (Reidel, Dordrecht, 1983).