Qualitative Spatial Reasoning in Interpreting Text and Narrative

Ernest Davis*
Dept. of Computer Science
New York University
davise@cs.nyu.edu

September 5, 2011

Abstract

Simple natural language texts and narratives often raise problems in commonsense spatial knowledge and reasoning of surprising logical complexity and geometric richness. In this paper, I consider a dozen short texts — five taken from literature, the remainder contrived as illustrations — and discuss the spatial reasoning involved in understanding them. I conclude by summarizing their common features, and by tentatively drawing some morals for research in this area.

Simple natural language texts and narratives often raise problems in commonsense spatial knowledge and reasoning of surprising logical complexity and geometric richness. In this paper, I consider a dozen short texts — five taken from literature, the remainder contrived as illustrations — and discuss the spatial reasoning involved in understanding them. The spatial issues that arise in these particular texts fall into three broad categories:

- Fitting objects into containers.
- Blocking, pursuing, escaping, and hiding.
- Manipulating household objects; in particular, buttons and cords.

I conclude by summarizing the common features of the spatial inferences involved here, and by tentatively drawing some morals for research in this area.

1 Seats in a bus

I begin with a passage from *Let Stalk Strine* by Afferbeck Lauder. The book is a satirical guide to the spoken Australian language; the passage, written in an Autralian accent, illustrates the use of the word "aorta"[1] ("they ought to"):

*This paper is an expanded version of a keynote speech given at COSIT 2011. Thanks to Max Egenhofer and Reinhard Moratz for the invitation to speak at COSIT; and to Antony Galton, Hector Levesque, and James Pustejovsky for helpful suggestions.

[1]Lauder's definition of "aorta" is, "the personification of the benevolently paternal welfare State to which all Strines — being fiercely independent and individualistic — appeal for help and comfort in moments of frustration and anguish".

Text 1.

> Aorta have more busses. An aorta mikem smaller so they don't take up half the road. An aorta put more seats innem so you doan tefter stann all the time. An aorta have more room innem - you carn tardly move innem air so crairded. Aorta do something about it.[2]

To understand this passage, a reader must realize that the changes that the putative speaker is calling for are mutually inconsistent, and infer from that that the speaker is a fool. It is certainly common even in unsophisticated narratives that a speaker's utterances are not to be taken at face value, but are primarily important for what they reveal about the speaker. However, in the context of reasoning systems, forward inference that a stated collection of constraints is inconsistent is a somewhat unusual inference direction.

Thus, the issue here for qualitative spatial reasoning (QSR) is the contradiction between increasing the size of each seat, increasing the number of seats, and decreasing the size of the bus. Let us approximate size as area — in this section, I will view things strictly in a two-dimensional setting — and approximate the bus in terms of its seating area, ignoring the rest of the bus. Then as a first stab at representation, we could use a simple rule such as

(1.1) $\texttt{AreaOf(Bus) = NumberOfSeatsIn(Bus) * AreaOf(Seat).}$

to infer that, if both the number of seats and the area per seat increase, then the area of the bus must increase. In an *ad hoc* notation — one can formalize this, but I will not bother —

(1.2) $\texttt{Area(Seat)}\uparrow \wedge \texttt{NumberOfSeatsInBox}\uparrow \Rightarrow \texttt{Area(Bus)}\uparrow$

The well-known theory of signed differential analysis (e.g. de Kleer and Brown 1985) allows us to infer (1.2) from (1.1).

However, not only has equation (1.1) been pulled out of thin air, it is also very narrow in scope; it applies only under the possibly unrealistic and certainly unnecessary assumption that all the seats have exactly the same area. What we really want to do is to use the general rule that, if you have a bunch of things, and you make each of the things bigger, and you add some more things to the bunch, then the bunch as a whole will occupy more space.

This involves two geometric rules:

(1.3) $[\forall_{\texttt{r}\in\texttt{c}}\ \texttt{Region(r)}] \wedge \forall_{\texttt{r1,r2}\in\texttt{c}}\ \texttt{DR(r1,r2)}] \Rightarrow$
$\texttt{AreaOf}(\bigcup \texttt{c}) = \sum_{\texttt{r}\in\texttt{c}} \texttt{AreaOf(r)}.$

(1.4) $\forall_{\texttt{r}}\ \texttt{Region(r)} \Rightarrow \texttt{AreaOf(r)} > 0.$

Rule (1.3) above states that, if `c` is a collection of non-overlapping region, then the area of the union of the regions in `c` is the sum of their areas. The predicate `DR(r1,r2)` is the RCC [12] relation, "Regions `r1` and `r2` do not overlap."

We can now characterize the required inferences as follows: There are two busses: `B1` the actual bus and `B2` the desired bus. Seats in a bus do not overlap. There is a one-to-one mapping (injection) `C` of the seats of `B1` into the seats of `B2`; each seat of `B1` is smaller than the matching seat in `B2`. The

[2] The passage is written in an Australian accent. Rewritten in standard English, it reads, "They ought to have more busses. And they ought to make them smaller so they don't take up half the road. And they ought to put more seats in them so you don't have to stand all the time. And they ought to have more room in them — you can hardly move in them, they're so crowded. They ought to do something about it." "Strine" is the Australian pronunciation of the word "Australian". The pen name "Afferbeck Lauder" derives from the phrase "alphabetical order".

region occupied by each bus is the union of the regions occupied by the seats plus a frame; and we will assume that the frame of B2 cannot be made smaller than the frame of B1. It then follows that B2 must be larger than B1.

At this point, let me make a disclaimer about use of logical notations in this paper that has been standard since at least Pat Hayes' "In Defense of Logic" of 1977 [5]. This first sections of this paper presents a lot of logical formulas. (In the later sections, the representational problems become too daunting.) I am not claiming that the proper design for a reasoning system is to manipulate these formulas or similar formulas symbolically. The purpose of the logical notation is just to serve as a standard notation in which to discuss knowledge and inference. I *am* claiming that this is a reasonable idealization or approximation of an inference that a reasoning program must somehow carry out, to understand these texts. I am also claiming, though with much less confidence, that such a reasoning system will have a declarative representation, at least of the problem specification, and probably of some substantial part of the background knowledge, which expresses something reasonably close to the formulas here.

That said, we can proceed with formalizing the inference as follows:

(1.5) $\texttt{Bus(b)} \wedge \texttt{s1} \in \texttt{SeatsOf(b)} \wedge \texttt{s2} \in \texttt{SeatsOf(b)} \wedge \texttt{s1} \neq \texttt{s2} \Rightarrow$
$\texttt{DR(PlaceOf(s1),PlaceOf(s2))}.$

(1.6) $\texttt{Bus(b)} \wedge \texttt{s} \in \texttt{SeatsOf(b)} \Rightarrow$
$\texttt{DR(PlaceOf(s),PlaceOf(FrameOf(b)))}.$

(1.7) $\texttt{Bus(b)} \Rightarrow$
$\texttt{PlaceOf(b) = PlaceOf(FrameOf(b))} \cup (\bigcup \{ \texttt{PlaceOf(s)} \mid \texttt{s} \in \texttt{SeatsOf(b)} \}).$

(1.8) $\texttt{Bus(B1)} \wedge \texttt{Bus(B2)}.$

(1.9) $\texttt{s} \in \texttt{SeatsOf(B1)} \Rightarrow$
$\texttt{C(s)} \in \texttt{SeatsOf(B2)} \wedge \texttt{AreaOf(PlaceOf(C(s)))} > \texttt{AreaOf(PlaceOf(s))}.$

(1.10) $\texttt{s1} \in \texttt{SeatsOf(B1)} \wedge \texttt{s2} \in \texttt{SeatsOf(B1)} \wedge \texttt{s1} \neq \texttt{s2} \Rightarrow$
$\texttt{C(s1)} \neq \texttt{C(s2)}.$

(1.11) $\texttt{AreaOf(PlaceOf(FrameOf(B2)))} \geq \texttt{AreaOf(PlaceOf(FrameOf(B1)))}.$

The conclusion $\texttt{AreaOf(PlaceOf(B2))} > \texttt{Area(PlaceOf(B1))}$ then follows these axioms plus basic axioms of set theory and addition.

I would presume that current state-of-the-art thereom provers can find this proof; though, since it involves a combination of arithmetic and set theory, it is not trivial. Certainly, it is an easy exercise for state-of-the-art proof verifiers. However, I do not know of any work in the qualitative spatial reasoning literature that has looked at this kind of inference.

Note that this formulation supports inferences that (1.1) and (1.2) do not; for example, the inference that, if you make some of the seats larger and leave the rest the same, you will need a larger bus.

Assuming there are finitely many seats, (1.3) can be proven by induction from the following two geometric statements plus the axioms of addition.

(1.12) $\texttt{DR(r1,r2)} \Rightarrow \texttt{AreaOf(r1} \cup \texttt{r2) = AreaOf(r1) + AreaOf(r2)}.$

and

(1.13) $\texttt{DR(r1,r3)} \wedge \texttt{DR(r2,r3)} \Rightarrow \texttt{DR(r1} \cup \texttt{r2,r3)}.$

1.13 is very much the kind of inference that has been looked at extensively in the QSR literature, particularly if we express union as a relation rather than a function:

(1.13') `DR(r1,r3)` $\wedge$ `DR(r2,r3)` $\wedge$ `Union(r1,r2,ru)` $\Rightarrow$ `DR(ru,r3)`.

So this is simply a composition rule, albeit one with three relations, one of them ternary, on the left side. I do not know if a composition-based calculus of RCC and Boolean relations that will support this has been worked out.

Formula (1.12) is more complex since it involves an RCC relation, a Boolean operator, and arithmetic. Formula (1.3) is more difficult yet, because it involves a finite collection of regions of indeterminate cardinality.

However, the fact that this inference involves second-order logic, which in general is horribly undecidable, does not mean that we would be justified in simply ignoring it. On the contrary: the text here is easily understood, and an understanding of the text clearly involves a collection of regions of unspecified cardinality. So there must be some way of dealing with this. I am not at all suggesting that we should be writing general purpose theorem-provers for second-order logic with arithmetic, geometry, and set theory. The inferences that are actually needed for commonsense reasoning, including this one, must surely lie in some fairly narrow, reasonably tractable, corner of that space. What we have to do is to find out what that corner is.

Going back to the busses, we may further note the important fact that the number of seats you can fit in a bus does not change as you drive the bus around town. (It is as important, and often as difficult, to be able to express frame axioms as to express axioms of change; it is a huge mistake to count on non-monotonic frame inferences to find the right thing to do in this kind of case until you have worked out yourself what the right thing to do is.) This inference involves bringing in rigid mappings and can be justified once we add the following rule.

(1.14) $\forall_{m,r1,r2}$ `RigidMapping(m)` $\wedge$ `Region(r1)` $\wedge$ `Region(r2)` $\Rightarrow$
`AreaOf(RigidImage(m,r1)) = AreaOf(r1)` $\wedge$
`RigidImage(m,r1` $\cup$ `r2) = RigidImage(m,r1)` $\cup$ `RigidImage(m,r2)` $\wedge$
`[DR(r1,r2)` $\Leftrightarrow$ `DR(RigidImage(r1),RigidImage(r2))]`.

2 Fitting

Our second text is a contrived one; it is an example from the paper "The Winograd schema challenge" by Hector Levesque [9]. A Winograd schema is a pair of sentences that differ in only one or two words and that contain an ambiguity that is resolved in opposite ways in the two sentences and requires the use of world knowledge and reasoning for its resolution. The schema takes its name from a well-known example by Terry Winograd [13]:

> The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.

If the word is "feared", then "they" presumably refers to the city council; if it is "advocated" then "they" presumably refers to the demonstrators.

Levesque proposes to assemble a corpus of such Winograd schemas that are

- easily disambiguated by the human reader (ideally, so easily that the reader does not even notice that there is an ambiguity);

- not solvable by simple techniques such as selectional restrictions
- not Googlable; that is, there is no obvious statistical test over text corpora that will reliably disambiguate these correctly.

The corpus would then be presented as a challenge for AI programs, along the lines of the Turing test. The strengths of the challenge are that it is clear-cut, in that the answer to each schema is a binary choice; vivid, in that it is obvious to non-experts that a program that fails to get the right answers clearly has serious gaps in its understanding; and difficult, in that it is far beyond the current state of the art.

A number of the schemas that Levesque presents in his paper as examples draw on spatial reasoning at least in part. One of the most interesting for our purposes is the following:

Text 2:

The trophy would not fit into the brown suitcase because it was too [small/large].

How does a reading system know that "it" refers to the suitcase if the last word is "small" and to the trophy if the last word is "large"?

The first task is to interpret the phrase, "because it was too large" in terms of its spatial content. It seems to me that the spatial content of these two sentences can reasonably be glossed as

> This trophy does not fit inside the suitcase, and no larger trophy fits inside the suitcase, but some smaller trophy does fit inside the suitcase.
> The trophy does not fit inside this suitcase or inside any smaller suitcase but it does fit inside some larger suitcase.

Symbolically,

(2.1) $\neg$`FitsIn(Trophy,Suitcase)` $\wedge$
[$\forall_{\mathtt{t}}$ `Larger(ShapeOf(t),ShapeOf(Trophy))` $\Rightarrow$ $\neg$`FitsIn(t,Suitcase)`] $\wedge$
[$\exists_{\mathtt{t}}$ `Larger(ShapeOf(Trophy),ShapeOf(t))` $\wedge$ `FitsIn(t,Suitcase).`]

(2.2) $\neg$`FitsIn(Trophy,Suitcase)` $\wedge$
[$\forall_{\mathtt{s}}$ `Smaller(s,Suitcase)` $\Rightarrow$ $\neg$`FitsIn(Trophy,s)`] $\wedge$
[$\exists_{\mathtt{s}}$ `Smaller(Suitcase,s)` $\wedge$ `FitsIn(Trophy,s).`]

Here `Trophy` denotes the object; the function `ShapeOf(t)` denotes the region occupied by object `t` in some standard configuration. The symbol `Suitcase` here refers to the inside of the suitcase; for simplicity, we take that to be a fixed region of space. The relation `FitsIn(o,r)` asserts that object `o` fits in region `r`.

In general, I am assuming that "α cannot ϕ because it is too θ" is to be interpreted as

$$\neg\phi(\alpha) \wedge$$
$$[\forall_{\mathtt{a}}\ \Theta(\mathtt{a},\alpha) \Rightarrow \neg\phi(\mathtt{a})] \wedge$$
$$[\exists_{\mathtt{a}}\ \Theta(\alpha,\mathtt{a}) \wedge \phi(\mathtt{a})]$$

where α is an entity, $\phi(\cdot)$ is a formula with one free variable, $\Theta(X,Y)$ is a comparator and θ is the unary characteristic corresponding to Θ. In sentence 1, α is the trophy, $\phi(X)$ is the formula, "X fits in the suitcase", θ is "large", and $\Theta(X,Y)$ is "X is larger than Y". In sentence 2, α is the suitcase, $\phi(X)$ is the formula, "the trophy fits in X", θ is "small", and $\Theta(X,Y)$ is "X is smaller than Y".

There is a three-way ambiguity in the sentence "The trophy does not fit in the suitcase" that is worth noting. The simplest meaning is that the trophy does not fit in the suitcase if the suitcase is otherwise empty. Alternatively, as spoken by a person who also has to pack clothes in the suitcase, it may mean that the trophy does not fit together with all the higher priority items that need to be packed in the suitcase. Finally, as spoken by a person who has already partially packed the suitcase, it may mean that the trophy cannot be fit into the suitcase without rearranging the items already packed. We will use the first of these here; the other two can be represented using the techniques discussed in section 3.

To describe the possible regions that an object `o` may occupy, we use the predicate `FeasibleShape(r,o)`, meaning that it is possible to manipulate `o` so that its spatial extent is exactly `r`. The nature of the class of regions satisfying this relation for a given object `o` depends on the physical characteristics[3] of `o`. If `o` is a rigid object, then all its feasible shapes are congruent without reflection; if `o` is more flexible, then a broader class of shapes is feasible. The standard shape of `o` is feasible for `o`.

(2.3) $\forall_{\texttt{o}}$ `FeasibleShape(ShapeOf(o),o).`

We can now define the relation `FitsIn(r,o)` as meaning that there is some feasible shape `s` of `o` that is a subset of `r`.

(2.4) `FitsIn(x,r)` $\equiv$
$\exists_{\texttt{s}}$ `FeasibleShape(s,o)` $\wedge$ `s` $\subset$ `r.`

Since the trophy is a rigid object, its feasible shapes are all rigid mappings of the standard shape.

(2.5) `FeasibleShape(r,Trophy)` $\equiv$
$\exists_{\texttt{m}}$ `RigidMapping(m)` $\wedge$ `r = RigidImage(m,ShapeOf(Trophy)).`

It is less clear what is the proper interpretation of `Smaller(a,b)` and `Larger(a,b)`. Some plausible candidates:

2.6a. `Smaller(a,b)` $\equiv$ `VolumeOf(a)` $<$ `VolumeOf(b)` .

2.6b. `Smaller(a,b)` $\equiv$ `DiameterOf(a)` $<$ `DiameterOf(b).`

2.6c. `Smaller(a,b)` $\equiv$ `a` $\subset$ `b.`

2.6d. `Smaller(a,b)` $\equiv$ $\exists_{\texttt{s}}$ `s` $>$ `1` $\wedge$ `b=Scale(a,s).`

Keep in mind that the texts do not directly compare the trophy to the suitcase; if we were obliged to find an interpretation justifying "The suitcase is smaller than the trophy", then certainly neither (2.4c) nor (2.6d) would be options. However, our text only compare the trophy to hypothetical trophies and the suitcase to hypothetical suitcases. so (2.6c) and (2.6d) may be reasonable.

All of these are problematic, in some respects. In all of them, the third conjuncts of (2.2) and (2.3) $\exists_{\texttt{t}}$ `Larger(Trophy,t)` $\wedge$ `FitsIn(t,Suitcase)` and $\exists_{\texttt{s}}$ `Smaller(Suitcase,s)` $\wedge$ `FitsIn(Trophy,s)` are vacuous; that is, true for all possible values of `Suitcase` and `Trophy`. Interpretations (a) and (b) seem overly restrictive; that is, there are cases where it seems reasonable to say "The trophy does not fit in the suitcase because it is too large" even though there actually exists an object of larger volume or diameter that does fit in the suitcase. With (c) by contrast, the second conjunct,

[3] It also depends on the class of "manipulations" that you are willing to consider, but we will ignore that here.

$\forall_t$ `Larger(ShapeOf(t),ShapeOf(Trophy))` $\Rightarrow \neg$`FitsIn(t,Suitcase)`

$\forall_s$ `Smaller(s,Suitcase)` $\Rightarrow \neg$`FitsIn(Trophy,s)`

are simply consequences of $\neg$`FitsIn(Trophy,Suitcase)` and add no additional information.

In this respect, interpretation (d), which at first glance seems like the least plausible interpretation of "smaller/larger", actually comes out best. If the suitcase is assumed to be convex, then, as with interpretation (c), the second conjunct is vacuous. However, if we consider a case that is not convex, such a case designed to fit a particular object, like a violin case, then it is possible that, though the trophy does not fit in the case, a scale expansion of the trophy would fit in the case, or, equivalently, the trophy would fit in a scale contraction of the case. Under that circumstance, it would, I think, seem odd to say that the trophy does not fit in the case because the trophy is too large, or because the case is too small. What you would say is that it doesn't fit because it is the wrong shape, or because the case is the wrong shape. So for interpretation (d) the second conjuncts of (2.2) and (2.3) are not vacuous; and in the cases where the second conjunct fails, it does indeed seem odd to say that the trophy does not fit because the trophy is too large or the case is too small. (In this case, interpretation (c) gives the wrong answer; interpretatations (a) and (b) give the right one.)

However, the converse does not hold; the restriction to scale expansions and contraction is very limiting. In the circumstance where the interior of the case is much larger in volume and diameter than the trophy, but neither the trophy nor any of its scale expansions fit in the case, then (2.1) under interpretation (2.6d) would accept this as "The trophy does not fit because it is too large," whereas I think clearly in this case as well it is better to say "The trophy does not fit because it is the wrong shape."

In any case, what is critical for the task of disambiguation is not so much that the correct interpretation should be strikingly reasonable as that the incorrect interpretation should be indisputably wrong. Here the issue is thankfully clear-cut. If we misinterpret "it" in the two sentences, we obtain

(2.7) $\neg$`FitsIn(Trophy,Suitcase)` $\wedge$
[$\forall_s$ `Larger(s,Suitcase)` $\Rightarrow \neg$`FitsIn(Trophy,s)`] $\wedge$
[$\exists_s$ `Larger(Suitcase,s)` $\wedge$ `FitsIn(Trophy,s)`].

(2.8) $\neg$`FitsIn(Trophy,Suitcase)` $\wedge$
[$\forall_t$ `Smaller(ShapeOf(t),ShapeOf(Trophy))` $\Rightarrow \neg$`FitsIn(t,Suitcase)`] $\wedge$
[$\exists_t$ `Smaller(ShapeOf(Trophy),ShapeOf(t))` $\wedge$ `FitsIn(t,Suitcase)`].

In both of these sentences, the second conjunct is in itself invalid — that is, false for all values of `Trophy` and `Suitcase` — under any of the proposed interpretations of `Smaller/Larger`. Therefore, the corresponding resolution of the anaphora can be rejected. The task of the spatial reasoner is to determine that fact, for whichever interpretation of `Smaller/Larger` is adopted. Equivalently, the reasoner has to determine that the negations of these are universally true; that is the two sentences, $\forall_{t,s}\exists_{t1}$ `Smaller(ShapeOf(t1),ShapeOf(t))` $\wedge$ `FitsIn(t1,s)` and $\forall_{s,t}\exists_{s1}$ `Larger(s1,s)` $\wedge$ `FitsIn(t,s1)` are true. One can debate what exactly it means for a trophy to be too large to fit in a suitcase, but a trophy really cannot be too small to fit in a suitcase.

The reader may reasonably object that the above discussion of the "correct" geometrical interpretations of "small" and "large" are misguided, as it is altogether unlikely that the speaker of this sentence had *any* well-defined geometric interpretation in mind, or even a well-defined logical formulation of "because it is too [small/large]." This objection, which is of course just one instance of the eternal misfit of language and logic, is certainly correct, but it seems to me that it only makes the problem more difficult. I believe that, in most cases, it is easier to work with a geometrically specific notion of "smaller" and "larger" than to try to characterize inference based on a geometrically indeterminate notion; and I certainly have no idea how one can analyze inferences based on logically indeterminate formulations.

3 Packing

Similar issues arise with greater logical complexity in the following text:

Text 3:

Each of my books fits in this box; however, they do not all fit in the box.

If I had enough boxes like this one, then I could pack all my books.

This is more difficult because there are now multiple books, and the second sentence refers to a hypothetical collection of boxes.

One representational difficulty here is the contrary-to-fact conditional in the second sentence, but we will pass over that, and focus just on the spatial issues. (My own feeling is that the formulation of this as a contrary-to-fact conditional is purely rhetorical, and that the actual content being expressed *is* the spatial proposition described below, rather than a statement about an alternative possible world in which I have more boxes.)

In the "trophy/suitcase" sentences, we simplified the representation of the suitcase by taking `Suitcase` to be a fixed region of space, corresponding to the inside of the suitcase. It seems clear that this simplification is not adequate here; an adequate representation must deal with the actual boxes, and not just their interiors.[4] We will assume that the objective is to put the books in the boxes and then to close the boxes, so that each book is inside the interior cavity formed by the box.

The analysis can proceed along the following lines. First, we define an "arrangement" of a set of solid objects:

Definition 1. *Let* `c` *be a set of solid objects. An* arrangement `a` *of* `c` *is a function mapping each object* `o` *in* `c` *to a feasible spatial region such that no two objects overlap.*

The predicate `Arrangement(a,c)` means that `a` is an arrangement of `c`. The function `PlaceOf(x,a)` denotes the region occupied by `x` in `a`. We thus have the axioms:

(3.1) $\mathtt{Arrangement(a,c)} \Rightarrow$
$\quad \forall_{\mathtt{x,y} \in \mathtt{c}}\ \mathtt{x} \neq \mathtt{y} \Rightarrow \mathtt{DR(PlaceOf(x,a),PlaceOf(y,a))}.$

(3.2) $\mathtt{Arrangement(a,c)} \land \mathtt{x} \in \mathtt{c} \Rightarrow \mathtt{FeasibleShape(PlaceOf(x,a),x)}.$

(3.3) $\mathtt{Arrangement(NullArrangement,}\emptyset\mathtt{)}.$

(3.4) $\mathtt{Arrangement(a,c)} \land \mathtt{x} \notin \mathtt{c} \land \mathtt{FeasibleShape(r,x)} \land$
$\quad [\forall_{\mathtt{y} \in \mathtt{c}}\ \mathtt{DR(PlaceOf(y,a),r)}] \Rightarrow$
$\quad \exists_{\mathtt{a1}}\ \mathtt{Arrangement(a1,c} \cup \{\mathtt{x}\}\mathtt{)} \land \mathtt{PlaceOf(x,a1)=r} \land$
$\quad [\forall_{\mathtt{y} \in \mathtt{c}}\ \mathtt{PlaceOf(y,a1) = PlaceOf(y,a)}].$

Axioms 3.3 and 3.4 guarantee that any way of placing finitely many objects in feasible, non-overlapping regions constitutes a valid arrangement. Axiom 3.3 asserts that the null arrangement is valid. Axiom 3.4 asserts that if you have an arrangement, you can extend it to a new object by leaving all the other objects where they are and putting the new object in any non-overlapping feasible region.

[4] In the talk delivered at COSIT, I used a representation based on the interior for simplicity.

For the purpose of packing, books can be viewed as rigid objects (assuming that they are not packed open!) i.e. the feasible shapes of a book are all congruent.

(3.5) $\texttt{Book(b)} \wedge \texttt{FeasibleShape(r1,b)} \wedge \texttt{FeasibleShape(r2,b)} \Rightarrow \texttt{Congruent(r1,r2)}$.

We will not here give an axiomatization of the regions that can be attained by a non-rigid box; this is complex and obviously varies with the type of box.

Two objects are geometrically identical if they have the same class of feasible shapes.

(3.6) $\texttt{GeomIdentical(x,y)} \equiv$
$\forall_{\texttt{r}}$ $\texttt{FeasibleShape(r,x)} \Leftrightarrow \texttt{FeasibleShape(r,y)}$.

The predicate `ClosedBox(o,i)` means that region `o` is a closed box with interior `i`; equivalently, region `i` is an interior cavity of region `o`. This can be defined in terms of the RCC relations and the Boolean operators:

(3.7) $\texttt{ClosedBox(o,i)} \Leftrightarrow \texttt{NTPP(i,o} \cup \texttt{i)} \wedge \texttt{EC(i,o)}$.

We next define the relation, "In arrangement `a`, object `b` is inside closed box `x`":

(3.8) $\texttt{InsideClosedBox(b,x,a)} \equiv \texttt{ClosedBox(PlaceOf(x,a),PlaceOf(b,a))}$.

We can now represent our text. Let `B` be a constant symbol representing the set of books, and let `X` be a constant symbol representing the box. "Each book fits inside the box" is represented,

(3.9) $\forall_{\texttt{b}\in\texttt{B}}\ \exists_{\texttt{a}}\ \texttt{Arrangement(a,}\{\texttt{b,X}\}\texttt{)} \wedge \texttt{InsideClosedBox(b,X,a)}$

"The books do not all fit inside the box," is represented,

(3.10) $\neg\exists_{\texttt{a}}\ \texttt{Arrangement(a,B} \cup \{\ \texttt{X}\ \}\texttt{)} \wedge \forall_{\texttt{b}\in\texttt{B}}\ \texttt{InsideClosedBox(b,X,a)}$.

"If I had enough boxes like this one, I could pack all my books into boxes," is represented

(3.11) $\exists_{\texttt{n}}\ \forall_{\texttt{c}}\ \texttt{CardinalityOf(c)=n} \wedge [\forall_{\texttt{x}\in\texttt{c}}\ \texttt{GeomIdentical(x,X)}] \Rightarrow$
$\exists_{\texttt{a}}\ \texttt{Arrangement(a,B} \cup \texttt{X)} \wedge$
$\forall_{\texttt{b}\in\texttt{B}}\exists_{\texttt{x}\in\texttt{c}}\ \texttt{InsideClosedBox(b,x,a)}$.

That is: There is a number `n` such that if `c` is a collection of `n` boxes identical to `X`, then there is an arrangement `a` that puts each of the books into a box.

The representation exhibits two forms of logical complexity. First there is the use of "arrangements," which are structurally complex. Second, formula 3.11 has four alternations of quantifiers. In complexity theory, alternation of quantifiers is generally the most recalcitrant form of complexity for first-order formulas; and four alternations is a lot. (One could argue that replacing $\forall_{\texttt{c}}$ in formula (3.11) by $\exists_{\texttt{c}}$ would give an equivalent statement with only two levels of alternation; if the statement is true of any set of geometrically identical boxes, then it is true of all sets of identical boxes. However, if you replace "geometric identical" by "geometrically similar" for some notion of similarity, then there is no alternative to using the universal quantifier.)

We omitted in this section the size comparisons "larger" and "smaller" that we dealt with in section 2. Our analysis in section 2 can be incorporated here by adopting the reasonable convention that, in loading boxes, "larger" and "smaller" refer to the *capacity* of the box, and not to the *walls* of the box.

4 Repacking

Our final text in the area of fitting things into a box is logically by far the most difficult.

Text 4.

> The equipment came out of the box, but now I can't fit it into the box.

This is (a) a common experience; (b) geometrically impossible; (c) not devoid of geometric content. For instance, one knows that this is more likely to happen if the objects are intricately shaped and were initially intricately intertwined than if they are all rectangular bricks; and that it is unlikely, though not impossible, that the objects were originally in a small box and now you can't even fit them into a large box.

I have no idea how to represent this.

One can imagine an intelligent system, whose knowledge of geometry is stronger than its knowledge of people or its faith in syntactic niceties, carrying out the following reasoning in trying to disambiguate "it" in the above sentence: It is impossible that the equipment does not fit into the box. On the other hand, it is certain that the box does not fit into itself. Therefore "it" must refer to the box.[5]

5 Blocking

We now turn to texts involving motion: blocking, hiding, pursuing, escaping.

Text 5.

> I tried to keep the dogs out of the kitchen by putting a chair in the middle of the doorway, but it was too wide.

Before proceeding with the spatial analysis, let me remark that this does not quite work as the basis for a Winograd schema. I had at one point thought of making this a Winograd schema by giving the alternatives [wide/small]. However, that does not work, because there is a selectional preference to associate "wide" with "doorway" and "small" with "chair". It is not a strict selectional constraint; one can certainly speak of a "wide chair" or a "small doorway" without seeming anomalous. But it is a strong enough preference that, if there is an ambiguity where this preference points in one direction, while a spatial or physical inference points in the opposite direction, the resulting text is confusing. For example in the sentenece

> ? I tried to keep the dogs out of the kitchen by putting a chair in the middle of the doorway, but it was too narrow

the reader may have difficulty realizing that, geometrically, "it" must refer to the chair, because "narrow" is a more suitable descriptor for a doorway than for a chair. This illustrates the very high bar that Winograd schemas must meet.

[5] In the early episodes of *Star Trek: The Next Generation*, Commander Data, the android, was unable to use contractions. This was so patently absurd, even by the standards of *Star Trek*, that it was later explained away as a deliberately engineered limitation, to make him seem less human, like his pale skin. I have often thought that if they had written the show so that he had trouble with reference resolution, that would have shown some insight. It would probably also have been unwatchable.

Let me assume that the dogs do not push the chair out of the way,[6] and therefore the chair will block them unless there is room for them to get easily by, on one side or the other.

We start with the representation of the statement that putting the chair in a particular position `PosC` will keep the dogs out of the kitchen. Since we are now doing spatio-temporal reasoning, we use *histories* [6]; that is, continuous functions from a time interval to the state of the world. We use the following primitives:

`StateOf(h,t,s)` — `s` is the state of the world at time `t` in history `h`.
`Start(h)` — the start time of history `h`.
`End(h)` — the end time of history `h`.
`ObjectsOf(h)` — the objects involved in history `h`.
`Fixed(x,r,h)` — Object `x` occupies region `r` throughout history `h`.
`Continuous(x,h)` — Object `x` moves continuously in `h`.

We posit that each state of history `h` is a valid arrangement of `ObjectsOf(h)` satisfying axioms 3.1—3.4 and that every object moves continuously in every history.

(5.1) `StateOf(h,t,s)` $\Rightarrow$ `Arrangement(s,ObjectsOf(h))`.

(5.2) `ObjectOf(o,h)` $\Rightarrow$ `Continuous(o,h)`.

We omit the definition of `Continuous`, which raises complex issues for region-valued fluents; see, for example, [4] and [1]. We can define `Fixed` as follows:

(5.3) `Fixed(x,r,h)` $\equiv$
  `ObjectOf(x,h)` $\wedge$
  $\forall_{t,s}$ `StateOf(h,t,s)` $\Rightarrow$ `r=PlaceOf(x,s)`.

The statement "Placing the chair at position `ChairPos` blocks the dogs out of the kitchen" can now be represented

(5.4) $\forall_h$ `Fixed(Walls,WallPos,h)` $\wedge$ `Fixed(Chair,ChairPos,h)` $\wedge$
    `MyDog(d)` $\wedge$ `ObjectOf(d,h)` $\wedge$ `DR(PlaceOf(d,Start(h)),Kitchen)` $\Rightarrow$
  `DR(PlaceOf(d,End(h)),Kitchen)`.

Here `Walls` are the architecture around the kitchen — walls, door frame, etc.; `WallPos` is the position of `Walls`.

The statement "There is no way to position the chair so as to block the dogs" is therefore

(5.5) $\neg$ [$\exists_p$ `FeasibleShape(p,Chair)` $\wedge$
    $\forall_h$ `Fixed(Walls,WallPos,h)` $\wedge$ `Fixed(Chair,p,h)` $\wedge$
      `MyDog(d)` $\wedge$ `ObjectOf(d,h)` $\wedge$
      `DR(PlaceOf(d,Start(h)),Kitchen)` $\Rightarrow$
    `DR(PlaceOf(d,End(h)),Kitchen)`].

[6]That depends on the particular dog. Among my own dogs, Jerry and Harmony, who were Airedales, were cautious about this kind of thing and would not push the chair, or even squeeze past it. Max, who was an Irish terrier, usually would not but might if he were particularly anxious to get past. With Jake, who was also an Airedale, but was very strange, it sufficed to put any token object in the doorway — a shoebox, say — and he would be convinced that he could not possibly get by on either side.

The logical form of the statement that the doorway is *too* wide follows our discussion in section 2; we will not write it out here, but note that it adds more complexity, and in particular another level of quantification.

Moreover, the reader of text 5 should understand that the reason that I put the chair in the center of the doorway is because I know that, if there is any position of the chair that blocks the dogs, then there is a central position of the chair that blocks the dog. To represent that, it will be convenient to define the predicate `Blocks(c,s,x,k)`, the collection of objects `c` as placed in state `s` block object `x` from entering region `k`.

(5.6) $\texttt{Blocks(c,s,x,k)} \equiv$
$\quad \forall_{\texttt{h}}\ [\forall_{\texttt{o}\in\texttt{c}}\ \texttt{Fixed(o,PlaceOf(o,s),h)}] \wedge$
$\quad\quad \texttt{DR(PlaceOf(x,Start(h)),k)} \Rightarrow$
$\quad \texttt{DR(PlaceOf(x,End(h)),k)}.$

We can now assert that, if it is possible to block the dogs out of the kitchen using chair `c`, it is possible to do so by placing the chair in the center of the doorway.

(5.7) $\forall_{\texttt{c}}\ \texttt{Chair(c)} \wedge$
$\quad [\exists_{\texttt{s}}\ \texttt{PlaceOf(Walls,s)=WallPos} \wedge$
$\quad\quad \forall_{\texttt{d}}\ \texttt{MyDog(d)} \Rightarrow \texttt{Blocks(\{Walls,c\},s,d,Kitchen)}] \Rightarrow$
$\quad [\exists_{\texttt{s1}}\ \texttt{PlaceOf(Walls,s1)=WallPos} \wedge \texttt{Central(PlaceOf(c,s1),Doorway)} \wedge$
$\quad\quad \forall_{\texttt{d}}\ \texttt{MyDog(d)} \Rightarrow \texttt{Blocks(\{Walls,c\},s1,d,Kitchen)}].$

Note the alternation of quantifiers and the quantification over both histories and arrangements. Note also the criticality of the continuity assumption in this reasoning; if the dogs could disapparate or teleport, then no chair would block them.

6 Hiding

We now turn to a narrative.

Text 6:

This is what those little rabbits saw round that corner!

Little Benjamin took one look, and then, in half a minute less than no time, he hid himself and Peter and the onions underneath a large basket ...

The cat got up and stretched herself, and came and sniffed at the basket.

Perhaps she liked the smell of onions!

Anyway, she sat down upon the top of the basket.

She sat there for *five hours.*

I cannot draw you a picture of Peter and Benjamin underneath the basket, because it was quite dark, and because[7] the smell of onions was fearful; it made Peter Rabbit and little Benjamin cry.

— *The Tale of Benjamin Bunny,* Beatrix Potter.

[7] This use of "because" is certainly remarkable.

Figure 1: From *The Tale of Benjamin Bunny*

Understanding this text and achieving narrative coherence involves the following inferences among others:

A. Initially, the cat does not see the rabbits because they are in back of her. (This is a text-image inference; you need the illustration (figure 1) in addition to the text.)

B. The cat cannot see the rabbits while they are underneath the basket.

C. The rabbits cannot get out from under the basket in its current position.

D. In order to get out, the rabbits must lift an edge of the basket.

E. If the basket is moved, the cat will be moved and will be disturbed.

Like the trophy/suitcase example discussed in section 2, these inferences have to do with a cavity enveloped by solid matter. Peter, Benjamin, and the onions are in the closed space formed by the basket and the ground; the cat is outside that space. Nothing can go from inside that space to outside that space without going through the solid, opaque material of either ground or basket. Rabbits cannot go through solid material; this is the justification of inference (C). Light cannot go through opaque material; this is the justification of inference (B) and the explanation of the fact that it is dark under the basket. (Finding explanations of explicit facts in terms of other explicit facts is important in narrative understanding, though harder to characterize than inferences. For one thing, if a fact has one explanation, one need not seek a different explanation.)

The inference for the material objects can be carried out using the following rules:

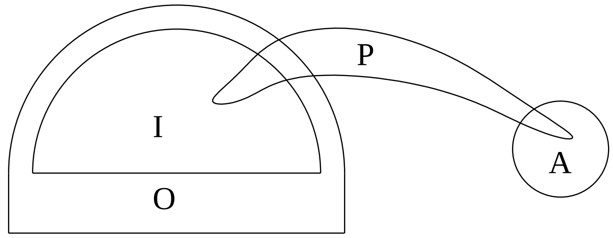

Figure 2: Rule 6.1

(6.1) $\texttt{ClosedBox(o,i)} \wedge \texttt{DR(a,o} \cup \texttt{i)} \wedge \texttt{C(p,a)} \wedge \texttt{C(p,i)} \wedge \texttt{Connected(p)} \Rightarrow \texttt{O(p,o)}$.

(6.2) $\texttt{O(x, y} \cup \texttt{z)} \Leftrightarrow \texttt{O(x,y)} \vee \texttt{O(y,z)}$.

Figure 2 illustrates rule 6.1. Here `i` is the region underneath the basket; `o` is the union of the basket and the ground; `a` is any region outside the basket; and `p` is a hypothetical path for anything (rabbits, cat) to go from outside the basket to underneath it, or vice versa. `O(x,y)` is the RCC relation "`x` overlaps `y`"; it is equivalent to $\neg$`DR(x,y)`.

To relate this to motion over time, we use axiom (6.3), which asserts that the swath of a moving object `x` during history `h` is a connected region of space ; and (6.4) which asserts that if the swath of `x` overlaps `r` then `x` itself overlaps `r` at some time between `t1` and `t2`.

(6.3) $\forall_{\texttt{h,x}}\ \texttt{x} \in \texttt{ObjectsOf(h)} \Rightarrow \texttt{Connected(Swath(x,h))}$.

(6.4) $\texttt{O(Swath(x,h),r)} \Rightarrow \exists_{\texttt{t,s}}\ \texttt{StateOf(h,t,s)} \wedge \texttt{O(PlaceOf(x,s),r)}$.

The problem specification is that the union of the ground and basket form a closed box; the rabbits are inside this box and the cat is outside.

(6.5) $\texttt{Arrangement(S0,}\{\ \texttt{Peter, Benjamin, Cat, Basket, Ground}\ \}\texttt{)}$.

(6.6) $\texttt{ClosedBox(PlaceOf(Basket,S0)} \cup \texttt{PlaceOf(Ground,S0) UnderBasket)}$.

(6.7) $\texttt{PlaceOf(Peter,S0)} \subset \texttt{UnderBasket} \wedge \texttt{PlaceOf(Benjamin,S0)} \subset \texttt{UnderBasket}$.

(6.8) $\texttt{DR(PlaceOf(Cat,S0), UnderBasket)}$.

The argument then proceeds as follows: First, applying rule (6.3), if the cat `x` goes from her position `c` outside the union of ground and basket `b` at time `t1` to make unfriendly contact with Peter `a` at time `t2`, then her swath `w` over `[t1,t2]` is a connected region that contains both her position at `t1` and her position at `t2`. Therefore `w` overlaps the union of ground and basket `b` (6.1); therefore `w` overlaps either the ground or the basket (6.2); therefore at some time `t` between `t1` and `t2` the cat either overlaps the ground or overlaps the basket (6.4). But two solid objects cannot overlap.

Analogous arguments show that the light cannot get into the basket, that the cat cannot see the rabbits under the basket, and that the rabbits cannot get out from under the basket.

Parts of this inference are within the scope of existing qualitative spatio-temporal reasoning systems; I do not know whether there is a reasoning system that can handle all of it.

6.1 Escaping by lifting the basket

Inference (C), by contrast, is not topological. If the cat were lying next to the basket rather than on top of it, then it might be possible (though certainly risky) to lift an edge of the basket without moving the cat. Likewise, if there was some non-rigid part of the rim — a door, or a cloth hanging, or such — then again the rabbits might be able to escape without moving the part of the basket in contact with the cat. (Another possible interpretation of the rabbits' predicament is that they can't lift the basket because the cat is too heavy.)

A common reaction here is to suppose that that the reader carries out the inference by *visualizing* or *simulating* the scenario; that is, the reader constructs a fully specified model of the starting state and uses physical simulation techniques to determine what will happen if the rabbits lift one edge or another of the basket (e.g. [7]). This is discussed further in section 10.1.

7 Fleeing

We continue with the theme of rabbits in jeopardy, this time from a dog.

Text 7.

> Many dogs, they say, are the death of a hare, a single dog cannot achieve it, even one much speedier and more enduring than Bashan. The hare can "double" and Bashan cannot — and that is all there is to it. . . . The hare gives a quick, easy, almost malicious twitch at right angles to the course and Bashan shoots past from his rear. Before he can stop, turn around, and get going in the other direction, the hare has gained so much ground that it is out of sight.
> — "A Man and his Dog", Thomas Mann

The text here provides some spatial information, but a great deal is left to the reader, who must make inferences such as

A. Before the double, Bashan and the hare are running in the same direction, the hare in front, and Bashan behind.

B. Assuming that Bashan is faster than the hare on a straight course, if the hare did not double, then Bashan would catch the hare.

C. At the time of the double, Bashan is quite close behind; perhaps a second or two. (The double doesn't accomplish anything if Bashan is 20 feet away.)

D. The distance of the jump in the double must be comparable to the width of Bashan (doubling an inch won't help); it is not large enough that a jump forward of that size would allow the hare to escape (the hare is not jumping 50 feet). Note that the word "twitch" here, which generally means a small muscular movement, is actually misleading; the reader must infer that it has to be taken figuratively.

E. A pack of dogs can catch the hare by surrounding it; the double wouldn't help.

Moreover, beyond these specific inferences, the reader can grasp how the double works. One aspect of this grasping is certainly that the reader can visualize the scenario. We will discuss the relation of inference and visualization in section 10.1.

At this point, and for most of the rest of the paper, I am abandoning my attempts at formal representation and am halting at informal discussion of the issues involved. The spatio-temporal theory needed is complex. The dog's pursuit and the hare's flight lie within the theory of differential games; since the hare uses continuous feedback from its perception of the dog and vice versa, it is necessary to reify control strategies for the hare and the dog, and to quantify over control strategies. The hare's advantage is that he can carry out an essentially discontinuous change of velocity, whereas the dog has a limit on the second derivative, at least when he is currently running.

Several of the inferences I have listed above are quantitative, in the sense of involving numerical values. Note, however, that these are all inferred – the text contains no numerical values and only one geometric or mathematical phrase "at right angles" — and that the reader can estimate them at best to within a factor of two or so.

8 Buttons

We now turn to texts dealing with household objects: Buttons in this section, cord in the next. The difficulties that arise in analyzing these inferences are formidable.[8] The spatial and spatio-temporal relations involved are complex; they are mostly implicit in the text; and they are hard to separate from the physical relations. The material physics itself is often difficult; it is not at all the kind of mechanics that one learns in freshman physics.

Let me start with three similar invented texts:

Text 8.

A. I tried to push the button through the hole, but it was too [large/small].

B. I tried to push the button through the hole, but it had been sewn too tightly to the coat.

C. I forgot that the top button was fastened, so when I took off the coat, it tore off.

Text 8.A is another Winograd schema. Linguistically, it is very similar to text 2 (the trophy/suitcase example), but the spatial issues are very different. What is impossible here is not the state of the button being in the fastened position, on the far side of the buttonhole — that may well be geometrically feasible — but the motion of getting the button through the hole. In that respect, it is analogous to the example of the dog getting into the kitchen, except that the motions involved in fastening a button are much more complex than the motions of a dog getting past a chair into a kitchen; you slide one side of the button through the hole, then you stretch the hole (i.e. the cloth around the hole) around the other side of the button, then you allow the hole to relax.

In text 8.B the physical issues involved in understanding the problem are even more difficult. Note also that if the final phrase "to the coat" is deleted, then the reference of "it" become genuinely ambiguous, with probably a preference for interpreting it as meaning the hole. The disambiguation in the full sentence depends on the fact that a button is sewn *to* a coat, whereas a buttonhole is certainly not sewn to a coat.[9]

Text 8.C deals with a common, accidental scenario with buttons. Note, first, that if we change

[8] I started work on buttons with Leora Morgenstern in 1985; we soon abandoned it as too difficult. Returning to the subject now it is disheartening to realize that all the work that has been done in the last 25 years does not bring us perceptibly closer to being able to deal with the problem.

[9] It is not clear what preposition *is* appropriate in the phrase "sew a buttonhole [·] a coat". My mother-in-law, Mavra Iano, who actually does sew buttonholes, says that the best choice is "on" but this seems awkward both to her and to me.

"fastened" to "unfastened", then the sentence becomes clearly anomalous; this indicates that the causal structure here is well understood.

A passage similar to 8.C appears in chapter 2 of *David Copperfield,* by Charles Dickens.

Text 8.D

> [Peggotty] gave [my head] a good squeeze. I know it was a good squeeze because, being very plump, whenever she made any little exertion after she was dressed, some of the buttons on the back of her gown flew off. And I recollect two bursting to the opposite side of the parlour while she was hugging me.
> — *David Copperfield,* Charles Dickens.

Achieving narrative coherence requires inferring that Peggotty's gown is very tight, and then constructing a causal chain incorporating her exertion, her plumpness, the tightness of the gown, and the bursting buttons. It is also, I think, relevant that, because the buttons are on the back of the gown, when they burst, David may be able to see them, but Peggotty herself certainly cannot.

9 Cord

Our final examples have to do with cords. First, another invented text.

Text 9.A.

> The power cord on the laptop would not reach from the desk to the outlet, so I got an extension cord. Then my wife objected to having an electric cord across the center of the living room ("Men!"), so I laid it around the edge of the room, and hid it behind the furniture, but of course I had to get a much longer cord.

Note that the disambiguation of "it" which refers, not to the extension cord previously mentioned but to the much longer cord mentioned later in the text. Note also that the events are narrated out of order; I got the longer cord before I laid it around the room and hid it behind the furniture.

The spatial features of cord used in this example are actually considerably simpler than the button examples. We can define a formal model of the kinematic properties of cord as follows: A piece of cord `c` has an associated length `LengthOf(c)` and an radius `RadiusOf(C)` which is much less than its length. In its standard configuration, the "core" of the cord, `CoreOf(c)` is the line segment along the x-axis from $\langle 0,0,0\rangle$ to $\langle \texttt{LengthOf(C)},0,0\rangle$. In that position, the cord fills the region `ShapeOf(C)`; this is a region that contains `CoreOf(c)` and lies inside the topologically open, right circular cone with axis `CoreOf(c)` and radius `RadiusOf(c)`.

A feasible configuration is determined by a function $\Gamma(X,Y,Z)$ mapping the standard configuration into 3-space. The function Γ has the following properties:

- Γ is a homeomorphism over `ShapeOf(c)`.
- Γ is twice differentiable over `CoreOf(c)`.
- Γ preserves arc-length over the core. That is, for any a,b such that $0 \leq a < b \leq \texttt{LengthOf(C)}$, Γ maps the line segment from $\langle a,0,0\rangle$ to $\langle b,0,0\rangle$ into a curve of arc-length $b-a$.
- The radius of curvature of $\Gamma(\texttt{CoreOf(C)})$ is everywhere at least `RadiusOf(C)` (possibly infinite).

- For any fixed value of x, the function $\lambda(y,z)\,\Gamma(x,y,z)$ is a rigid mapping. That is, Γ preserves the shapes of the cross-sections.

This allows the cord to have a non-constant cross section and to bend and twist, while perserving constant length and constant cross-sections and preventing it from overlapping itself.

Given this ontology, one can represent statements like, "All the feasible positions for the first extension cord that reach from the outlet to the table go through the middle of the living room," and "There are feasible positions for the second extension cord that go from the outlet to the table, but remain at the edge of the room." Let `C1` be the cord for the laptop itself; `C2` be the combination of `C1` with the first extension cord; and `C3` be the combination of `C1` with the second extension cord. Assume for simplicity that `Laptop` and `Outlet` are fixed regions of space. Define the predicate `Reaches(c,x,y,a)` meaning that cord `c` reaches from region `x` to region `y` in arrangement `a`.

(9.1) $\mathtt{Reaches(c,x,y,a)} \equiv$
$\mathtt{At(PlaceOf(End1(c),a),x)} \wedge \mathtt{At(PlaceOf(End2(c),a),y)}$.

(9.2) $\mathtt{Cord(C1)} \wedge \mathtt{Cord(C2)} \wedge \mathtt{Cord(C3)}$.

(9.3) $\neg\exists_{\mathtt{a}}\ \mathtt{Reaches(C1,Laptop,Outlet,a)}$.

(9.4) $\mathtt{Reaches(C2,Laptop,Outlet,S0)} \wedge \mathtt{Through(PlaceOf(C2,S0),LivingRoom)}$.

(9.5) $\forall_{\mathtt{a}}\ \mathtt{Arrangement(a,\{C2\})} \wedge \mathtt{Reaches(C2,Laptop,Outlet,a)} \Rightarrow$
$\mathtt{Through(PlaceOf(C2,a),LivingRoom)}$.

(9.6) $\mathtt{Reaches(C3,Laptop,Outlet,S0)} \wedge \neg\mathtt{Through(PlaceOf(C3,S0),LivingRoom)}$.

`End1(c)` and `End2(c)` are *pseudo-objects* [2]; that is, geometric entities that "move around with" an object.

The final text is the account of the death of the villain Bill Sykes in *Oliver Twist*. Sykes is trying to escape from an angry mob. He is on the roof of a building and plans to lower himself to the ground with a rope.

Text 9.B.

[H]e set his foot against the stack of chimneys, fastened one end of the rope tightly and firmly around it, and with the other made a strong running noose . He could let himself down by the cord to within a less distance of the ground than his own height, and had his knife ready in his hand to cut it then and drop.

At the very instant when he brought the loop over his head previous to slipping it beneath his arm-pits . . . at that very instant, the murderer, looking behind him on the roof, threw his arms above his head, and uttered a yell of terror.

"The eyes again!" he cried in an unearthly screech.

Staggered as if by lightening, he lost his balance and tumbled over the parapet. The noose was on his neck. It ran up with his weight, tight as a bowstring, and swift as the arrow it speeds. He fell for five-and-thirty feet. There was a sudden jerk, a terrific convulsion of the limbs; and there he hung, with the open knife clenched in his stiffening hand.

The old chimney quivered with the shock, but stood it bravely. The murderer swung lifeless against the wall.

— *Oliver Twist,* Charles Dickens.

This rather lurid passage is a treasure trove of fascinating, very difficult, commonsense inferences. The reader must understand, certainly both what actually happened and what Sykes had in mind, and also, I would argue, what would have happened under a number of alternative scenarios. This involves answering questions such as:

- What is the actual cause of Sykes' death? (It is not explicit.)
- How is he hanging at the end?
- Why does the chimney quiver?
- What would have happened if Sykes had stumbled:
  - Before attaching the rope to the chimneys?
  - Before slipping the noose over his head?
  - After slipping the noose under his armpits?
  - If the chimney(s) had broken under the shock?

There are also, of course, interesting inferences here that are not spatial and physical, most obviously why Sykes hallucinates seeing Nancy's eyes and why this causes him to stumble.

10 Summary

As a source for interesting problems in qualitative spatial reasoning, the interpretation of natural language text has some advantages over alternative sources, such as problems in manipulation or planning:

First, the problems are natural and interesting. By contrast, qualitative spatial reasoning tasks that are generated purely by considering the mathematical structure of simple geometric relations are generally artificial and all too often of purely theoretical interest.

Second, it is clear that the problems are indeed qualitative, and that there is no alternative to carrying out spatial inferences based on partial information. By contrast, in applications such as robotic manipulation and planning, or game playing applications, the temptation is always to reduce the role of qualitative reasoning as far as possible by aggressively collecting or positing as much precise information as possible.

Third, as we have seen, text interpretation involves a wide range of spatial relations, logical forms, and direction of inference. Other applications often focus on problems of greater size but with much fewer variations in structure.

Fourth, the reductive reactions, "You should just use simulation" or "You should just use corpus-based learning" are much less plausible suggestions for how to deal with qualitative reasoning in the context of text understanding than in other applications. Simulation and machine learning have made extraordinary advances in the last decade, and there is a strong temptation to turn to one or the other of these techniques as a panacea for intelligent tasks of all kinds. (Automated reasoning has also made great advances, but in a more narrow range of applications, such as program verification, that are less visible to the general public, or even the AI community.) This role that these techniques might play in this kind of reasoning is a large issue; we will return to this point below.

On the other hand, natural language interpretation also has serious drawbacks as an application for qualitative spatial reasoning. One is that interesting spatial reasoning problems arise very erratically. The quotes from novels and short stories in this paper are not representative of narrative text generally. You can read *Pride and Prejudice* from one end to the other and hardly find a non-trivial spatial inference; from that standpoint the characters might practically as well be disembodied spirits. The same is true, many days, of the newspaper. Generating spatial reasoning problems by inventing rather than finding texts is obviously problematic in many ways.

One might suppose that a way around this would be to use more technical materials, and indeed I have been looking at laboratory assignments for freshman chemistry with that in mind. However, I have not found this to be very productive. The direction of inference is heavily biased toward physical prediction. The spatial relations tend to be either very simple relation such as `In(A,B)`; or implicit, along the lines of the button example; or precise rather than qualitative. These limitations apply even more strongly to textual descriptions of mechanical devices. It is certainly possible that there is a category of texts that yields frequent, interesting, qualitative spatial inferences; but I have not found it.

A more profound problem is that, though textual understanding generates interesting individual problems, they are very haphazard in form. It does not generate any systematic class of problems. For both the theory and practice of computer science, this is very problematic. A computer program solves a class of problems that can be described systematically and is generally in principle infinite or at least very large. The only obvious class of problems that contains all the inferences we've looked at in this paper is the class of all inferences over the complete theory of geometry with set theory, which is a far broader category than we either need or can deal with.

There are other difficulties as well. Reasoning about shapes of objects is often critical in interpreting these texts, but the information about the shape that is used is, more often than not, implicit in the category of object rather than explicit in the text. Natural language is notoriously poor for shape description. For example, understanding either Bill Sykes' fate or his intended plan requires some degree of reasoning about the shape and functional structure of the human body — if he were spherical, or if he had been hung by his foot, the consequences would have been different — but it is not clear what information, or how much information, or how precise information is needed; and the text itself, of course, gives no help in resolving this issue. More generally, the process of translation from text to geometry is difficult. As we remarked above, in representing the statement "The trophy is too large for the suitcase", it is not obvious what precise geometric meaning to assign to "large" and what precise logical meaning to assign to "too"; and it seems all too likely that human natural language users do not actually have precise meanings in mind at all. Finally, in many cases, such as the button, it is hard to know where to draw the line between spatio-temporal reasoning on the one hand and physical or other domain reasoning on the other. Perhaps no such line should be drawn; but then it is much harder to see how to structure theory development.

10.1 Simulation

It is particularly tempting to suggest the use of simulation for all kinds of spatial and physical reasoning tasks, for two reasons. The first, as mentioned above, that simulation has become an extremely powerful technique in the last few years, in both the scientific computation community and, in a somewhat different form, in the animation community. (The scientific computing people emphasize accuracy and complexity; the animation people emphasize visual plausibility and real-time speed.) For example Rahimian et al. [11] describe carrying out simulations of 200,000,000 deformable red blood cells floating in plasma. With that kind of power available, if it is in any way possible to reduce an inference problem to simulation, then one can afford a very large degree of inefficiency in the reduction. An axe is not the ideal tool for opening the front door; but if axes are easily available and you can't figure out how to manufacture a key, and you need to get past the door, then the axe will get you there.

The second is that people faced with spatial and physical reasoning tasks often visualize the scenario; and visualization seems very similar to simulation. To what extent the cognitive processes involved in visualization actually resemble simulation is much debated. Let me not enter into that, and accept, at least for argument's sake, that visualizing a scenario is the mind's way of carrying out a simulation.

However, I would argue that the proposal to reduce spatial and physical reasoning to simulation is less appealing for the tasks of text and narrative interpretation discussed in this paper than for many other reasoning tasks, for a number of reasons.

Fundamentally, simulation carries out prediction for problems with precise physical and geometric specifications. Now, there are techniques for adapting simulation to use with qualitative specifications, of which the most obvious and the most general is to do Monte Carlo search within the space of scenarios that meet the qualitative specifications. This of course will tend to miss scenarios that are unlikely but not impossible; depending on the application, this can be a bug or a feature. It also delivers estimates of the probabilities of possible inferences; these can be useful. However, as the number of underspecified entities, their structural complexity, and the complexity of the constraints that form boundary conditions to the problem increase, the difficulty of carrying out the simulation and getting useful answers increases rapidly, and the meaningfulness of the probabilistic estimates obtained decreases. Doing Monte Carlo simulation over a collection of numerical parameters that lie in numerical intervals is one thing; doing Monte Carlo simulation over a collection of underdetermined regions and paths, subject to a collection of constraints on topological relations, surface normals, and so on, is a much more problematic undertaking. (To be fair, this kind of complexity make difficulties for *any* known reasoning technique; but the point is that the seeming simplicity of simulation vanishes.)

Simulation can also be adapted, in a number of ways, to inference directions other than prediction. But again the costs in terms of efficiency are large especially when structurally complex entities and deep levels of quantifier alternation are involved.

The features that we have noted in natural language interpretation — the multiple directions of inference, the logical complexity, the ubiquity of qualitative information, the heterogeneous geometrical constraints — thus play to the weak points of simulation-based techniques and not to its strengths. All this does not *prove* that simulation may not be a workable technique or even the best technique. But it does suggest strongly that it will not be at all easy to get it to work.

As for visualization: Taking my own introspections as the sole empirical evidence — being, of course, fully aware of the unreliability, not to say worthlessness, of conclusions drawn on that basis — it seems to me that, in encountering some of these texts, particularly text 7 with the dog chasing the hare, visualization is strongly involved; with others, such as the trophy fitting in the suitcase, I can easily understand it with no need to visualize, and visualization comes only as an afterthought, if I want to mull over the passage. However, from the standpoint of designing an automated system that will use simulation in the way that a human reader uses visualization, the critical question is, are the inferences that we have discussed above the *consequences* of simulations, or are they *inputs* guiding simulation? For instance, in the "dog chasing rabbit" example, there is the inference that the distance of the jump in the double is more than a few inches. Is that inference made by made by simulating a jump of a few inches and observing that it doesn't work? Or the inference made first, as a constraint on simulating the scenario? I don't know. But certainly the introspective experience of visualization does not suggest that I am visualizing a range of scenarios; it *seems* that I am just visualizing the one, correct scenario. *If* this is a correct reflection of the cognitive processes, and *if* we want to model the role of simulation in AI programs on the role of visualization in human understanding, then the inference has to be carried out prior to the simulation.

The "trophy in the suitcase" example is similar. Possibly one visualizes a trophy not fitting into a suitcase. But to justify the conclusion that the trophy is "too large" to fit into the suitcase, one would have to visualize a smaller trophy that does fit in the suitcase, and enough larger trophies that do not fit to justify the generalization that no larger trophies fit in the suitcase. (Another possibility, perhaps, it that you can visualize the trophy not fitting into the suitcase because it is too large. But I do not have any idea what that means, or what it would mean in terms of simulations.)

10.2 Corpus-Based Machine Learning

The successes of machine learning are, of course, even more astonishing and multifarious; and certainly in the long run a complete AI system would learn all this kind of knowledge. But at the current state of art, this seems like an even less promising approach. While the difficulty of finding sample texts that generate systematic sets of problems is certainly a serious obstacle for any approach to reasoning, it is an almost insuperable obstacle for corpus-based ML. The many successes of applications of corpus-based ML to natural language text are very explicitly based on the avoidance of the kind of inferences discussed here.

11 Morals for research

I should like to make a few tentative suggestions for research in this direction. Let me preface my suggestions, however, with the comment that these are not at all intended as *career* advice; and in particular, I strongly urge readers who are in pursuit either of a degree, or of tenure, or of government funding, at least in the US, not to put very many of their eggs into these baskets.[10] I hereby disclaim any responsibility, either moral or legal, for adverse effects on the career of anyone who takes the suggestions below seriously.

First, it may be worthwhile to take a number of carefully selected texts, such as the passage from *Oliver Twist* above, and do systematic, deep studies of the spatial and physical reasoning involved in a deep understanding, comparable to the large body of work on analyzing the plans, goals, interagent relations and so on in narrative text (e.g. [3]). (Of course, the issues of plans and goals are generally more central to the narrative structure than spatial issues.) Once a large enough corpus of such texts have been analyzed, some systematic patterns, and perhaps even a well-defined problem class, may emerge.

Second, it may be worthwhile to study the problems in carrying out inferences involving small, logically and geometrically complex systems of relations, rather than large systems of simple relations. Even if one supposes that the number of such inferences that correspond to "easy" reasoning in reading is finite and not very large and thus can be manually coded (which seems unlikely to me), the task of actually formulating such a collection of inferences and and of retrieving the needed inference when needed is a major undertaking.

Third, since we do not know how to evaluate an inference engine for small heterogeneous inferences, such as I am proposing, I suggest that we table the issue of evaluation until we understand the task better. Past a certain point, evaluating on an inappropriate standard is more damaging than failing to evaluate; this is notorious in education, but I think we have seen instances of it in AI as well.

Fourth: One way to deal with the logical complexity described in this paper is by "chunking"; defining a library of relations each of which expresses an important, complex relation. For example,

[10] I apply this caution to my own students. I currently have two students who are working on master's theses with spatial aspects. Azam Asl is working on the qualitative calculus of three-dimensional rotations, and Emily Morton-Owens is working on a spatio-temporal database for biographical information, created by information extraction from Wikipedia. I would certainly not recommend to either of them, or to any other student, that they work on this kind of issue instead. As for funding, I have almost given up trying to get it.

one might have the object level relation `Blocks(c,s,x,k)` meaning "In situation `s` obstacle `c` blocks `x` from entering region `k`" or the higher level relation "Too α to ϕ." This is analogous to the use of of a library of standard functions to simplify the problems of software development. I can hardly doubt that this is the correct approach; nonetheless, it is worth pointing out a couple of the difficulties that are likely to arise. (The analogous difficulties are common in software libraries.)

- Complex relations such as these tend to appear in many variants, with slightly different definitions and sets of arguments.
- A perennial problem with formal software specifications is that it is extremely difficult to fully specify the desired behavior of a library function of any complexity. In other words, though the routine is a useful chunking mechanism for the human programmer, it is hard to get the automated verifier to make any use of that chunking, and to reason at a high-level about the about the subroutine without looking each time at its internal code. The analogous problem here is that it is very difficult to formulate a set of high-level inference rules for these relations that will enable an automated reasoner to carry out the inferences you want, without descending into specifics of their definition in terms of low-level primitives.

Finally: I do not know to what extent the problems described here would interest linguists and cognitive psychologists, but it seems to me that we in AI would certainly benefit from some advice here as to which of these sentences actually need spatial reasoning, what kinds of inferences people seem to be carrying out, and so on.

12 Final remark

We have been studying automated commonsense reasoning for AI for many years. McCarthy published "Programs with Common Sense" [10] more than fifty years ago; work on CYC [8] was begun twenty-five years ago. As far as I can tell, we are still nowhere near being able to implement the kind of reasoning needed for the understanding of these simple texts. I do not see that we can predict with any confidence that another fifty years will bring us substantially closer. It seems to me, therefore, that, though the subject is certainly worth continuing to pursue, it should be viewed and presented as a high-risk, high-payoff enterprise, comparable to nuclear fusion reactors or SETI (search for extra-terrestrial intelligence).

However, for my taste, it is much more fun than either of those.

References

[1] E. Davis, to appear. Qualitative Reasoning and Spatio-Temporal Continuity. In S. Hazarika (ed.) *Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions.*

[2] E. Davis, 1988. A Logical Framework for Commonsense Predictions of Solid Object Behavior. *AI in Engineering,* vol. 3 no. 3, pp. 125-140

[3] M. Dyer, 1983. *In-depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension,* MIT Press.

[4] A. Galton, 2001. *Qualitative Spatial Change,* Oxford University Press.

[5] P. Hayes, 1977. In defense of logic. *IJCAI-77*, 559-565.

[6] P. Hayes, 1979. The naive physics manifesto. In *Expert Systems in the Micro-electronic Age,* ed. D. Michie, Edinburgh: Edinburgh U. Press.

[7] B. Johnston and M.A. Williams, 2007. A generic framework for approximate simulation in commonsense reasoning systems, *Commonsense-07.*

[8] D. Lenat, M. Prakash, and M. Shepherd, 1986. CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks, *AI Magazine,* Vol. 6, No. 4, 65-85.

[9] H. Levesque, 2011. The Winograd schema challenge, *Commonsense-11.*

[10] J. McCarthy, 1959. Programs with common sense, *Proc. Symposium on Mechanization of Thought Processes, I.*

[11] A. Rahimian et al. 2010, Petascale direct numerical simulation of blook flow on 200K cores and heterogeneous architectures. *SuperComputing-10.*

[12] D. Randell, Z. Cui, and A.G. Cohn, 1992. A spatial logic based on regions and connection, *KR-92*, 165-176.

[13] T. Winograd, 1972. *Understanding Natural Language,* Academic Press,