Observer Relativity, Physical Properties and Computation

Graham White [1]

Abstract. We examine Searle's argument, in [14], that syntax, and *a fortiori* computation, are not genuine physical properties of objects such as computers or minds, but are essentially observer-relative properties. He bases this argument on a sharp distinction between brute physical properties and observer-relative properties. We argue that what is at issue in Searle's argument is a certain intensionality of explanatory language: we give examples, and argue, on the basis of those examples, that this intensionality does not warrant the leap to observer-relativity.

1 Searle's Distinction

Searle [14, Ch. 9] has an extended and complex critique of cognitive science, one strand of which is the contention that syntax is not a *bona fide* physical concept: he argues for this by distinguishing between concepts which intrinsically apply to brute matter and those which are in some way observer-relative.

> The aim of natural science is to discover and characterise features that are intrinsic to the natural world. By its own definitions of computation and cognition, there is no way that computational cognitive science could ever be a natural science, because computation is not an intrinsic feature of the world. It is assigned relative to observers. [14, Ch. 9 §V]

Consequently, computational interpretations of the brain do not, strictly speaking, stand for anything:

> The thesis [of standard cognitive science] is that there are a whole lot of symbols being manipulated by the brain, 0's and 1's flashing through the brain at lightning speed and invisible not only to the naked eye but even to the most powerful electron microscope, and it is these that cause cognition. But the difficulty is that the 0's and 1's as such have no causal powers because they do not even exist except in the eyes of the beholder. The implemented program has no causal powers other than those of the implementing medium because the program has no real existence, no ontology, beyond that of the implementing medium. Physically speaking, there is not such thing as a separate "program level". [14, Ch. 9 §VII]

And so cognitive science – thus Searle – relies on a tacit appeal to a homunculus every time it ascribes computation to objects such as computers, or to the brain considered as a physical object, since

> in the commercial computer the ascription is always observer relative, the ascription is made relative to a homunculus who ascribes computational interpretations to the hardware states. Without the homunculus, there is no computation, just an electronic circuit. . . . [W]ithout a homunculus, there is no explanatory power to the postulation of the program states. There is just a physical mechanism, the brain, with its various real physical and physical/mental causal levels of description. [14, Ch. 9 §VII]

1.1 Searle's Arguments

One of Searle's arguments for this position, and against the orthodox position that the brain implements mental process, is to do with the properties of the relation "— implements —". Firstly, "we could make a system that does just what the brain does out of pretty much anything" [14, Ch. 9 §IV], and secondly that

> On the standard textbook definition of computation, it is hard to see how to avoid the following results:
>
> 1. For any object there is some description of that object such that under that description the object is a digital computer.
> 2. For any program and for any sufficiently complex object, there is some description of the object under which it is implementing the program. Thus for example the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecule movements which is isomorphic with the formal structure of Wordstar. But if the wall is implementing Wordstar, then if it is a big enough wall then it is implementing any program, including any program implemented in the brain [14, Ch. 9 §V]

Now, as he points out later, these "results" are not serious possibilities:

> I think it is possible to block the result of universal realisability by tightening up our definition of computation. . . . a more realistic definition of computation will emphasise such features as the causal relations among program states, programmability and controllability of the mechanism, and situatedness in the real world. . . . [T]here must be a causal structure sufficient to warrant counterfactuals. [14, Ch. 9 §V]

But, despite these qualifications, the serious problem is the many-to-many nature of the implementation relation, which shows that computation cannot be an intrinsic property of a physical object:

> [T]he really deep problem is that syntax is essential an observer-relative notion. The multiple realisability of computationally equivalent processes in different physical media is not

---

[1] Electronic Engineering and Computer Science, Queen Mary, University of London, email: g.graham.white@gmail.com, url: http://www.eecs.qmul.ac.uk/∼graham

> just a sign that the processes are abstract, but that they are not intrinsic to the system at all. They depend on an interpretation from outside. [14, Ch. 9 §V]

So here we have our basic contention: physical properties are intrinsic, whereas computational properties are not. This is what the argument seems to come down to.

1.2 Evaluation

This argument is complex and rather difficult to summarise; it has not been received entirely positively. Bringsjord and Noel regard the Chinese Room Argument as "arguably the twentieth centuries' greatest philosophical polarizer" [3, p. 144], and Searle's arguments around observer-relativity seem to be of the same nature. In particular, Endicott [4] concludes a critical review of Searle's arguments with the verdict that "Searle has failed to generate any convincing argument against cognitivism".

One of the main problems seems to be this. Searle has a distinction between observer-relative and non-observer-relative notions, or, as he elsewhere describes it, between extrinsic and intrinsic descriptions of a particular system. His problem about multiple realisability seems to fit in here: a realisation of a computation in a particular physical object will be associated with a set of concepts, namely those which describe the object in terms of its role in the computation. And these concepts can be thought of as part of a particular observer's view of the physical object, the view which describes it as performing a particular computation. By contrast, there will be other descriptions – those in terms of "physical features", in Searle's terms – which are not thus observer-relative. So there is a distinction between particular vocabularies – the observer-relative and non-observer-relative ones – and, regardless of the metaphysics, one can ask purely *logical* and *semantic* questions about them. Given a descriptive vocabulary, or a descriptive language using that vocabulary, can one tell just by looking it, without using question-begging terms like "computational", whether it is observer-relative or non-observer-relative? What is the relation (logical or semantic) between the observer-relative and non-observer-relative languages?

These questions turn out to be difficult to answer, both in general and also in the case of Searle's particular distinction between observer-relative and non-observer- relative. Endicott has a helpful analysis of Searle's position here:

> Searle has an exasperating proliferation of meanings for the term "intrinsic". [at one point] it means (a) *determination* by a set of lower-level properties, i.e. syntax is not intrinsic to physics because physics will not suffice to determine syntactic properties. But elsewhere it means (b) *definability* in terms of a particular class of predicates, i.e., syntax is not intrinsic to physics because it is "not defined in terms of physical features". But more often than not, Searle uses "intrinsic" to mean (c) *real*, ontologically speaking, and thus [to] mark the distinction between "the real thing" as opposed to the merely "as if", "derived", or "observer-relative". (Endicott [4, p 102 n3])

1.2.1 Searle's Languages

Searle's own practice seems to be more or less like this. He views the physical level of description – he himself describes this level in terms of "particles and fields" – as "completely objective" [14, Ch. 4 §II]. There is a hierarchy of more complex entities – systems, organisms – built of particles and fields [14, Ch 4 §II]. This pattern is quite pervasive:

> This, then, is one of the chief lessons of atomic theory: many features of big things are explained by the behaviour of little things. [14, Ch. 4 §1]

There is a sort of emergence here: the features of the "big things" are consequences of facts about the "little things", together with causal properties of those little things. Searle calls high-level features of this form "causally emergent system features" [14, Ch 5 §I].

Consciousness, too, is a causally emergent system feature [14, Ch 5 §I]: it is "not a 'stuff', it is a feature or property of the brain" [14, Ch 4 §III]. With consciousness comes subjectivity:

> the world itself has no point of view, but access to the world through my conscious states is always perspectival, always from my point of view. [14, Ch 4 §II]

Despite being causally emergent, consciousness cannot be reduced to mere properties of the brain [14, Ch 5 §II.5], because of the phenomenological irreducibility of conscious experience and because of the sensitivity of our definitional practices to that irreducibility [14, Ch 5 §IV].

This basic ontology (systems of particles plus causal relations) also plays a role in what Searle calls "multiple realisability":

> For any program and for any sufficiently complex object, there is some description of the object under which it is implementing the program. [14, Ch 9 §V]

The description would, presumably, consist of an assignment of, for example, components of a suitable Turing machine[2] to particular collections of particles of the object such that the causal relations between those particles amounted to an implementation of the said Turing machine.

1.2.2 Analysis

We can translate this position on multiple realisability into the following logical formalism. Given a description of an object X under which X implements program P, then we can concoct a singular term a_P which describes X as a particular object which contains such and such particles with such and such causal relations between them. Similarly, given another program P', we can find a singular term $a_{P'}$ which also refers to X but which does so by way of talking of the constituents of X by means of which X implements P'. So we have

$$a_P \quad = \quad a_{P'}, \tag{1}$$

and it is analytic that

$$a_P \quad \text{implements} \quad P \tag{2}$$

and it is also analytic that

$$a_{P'} \quad \text{implements} \quad P'; \tag{3}$$

---

[2] Searle clearly thinks that computation is essentially computation by Turing machines [14, Ch 9 §IV].

however (because it will not in general be the case that either a_P or $a_{P'}$ is a complete description of X), it will in general neither be analytic that

$$a_P \quad \text{implements} \quad P' \tag{4}$$

nor will it be analytic that

$$a_{P'} \quad \text{implements} \quad P \tag{5}$$

So the terms a_P and $a_{P'}$ are coreferential but have different senses: that is, they are *intensionally* different. Furthermore, propositions about implementation involve what philosophers of language *opaque contexts*, that is, operators applied to terms which were sensitive to the sense, not merely the reference, of their arguments. These claims would seem to be uncontentiously a consequence of Searle's claims about multiple realisability.

We should note that Searle takes his argument to be an argument about mental capacities, and therefore (because he takes intentionality-with-a-t, that is, the ability to refer, to be a feature of the mental) about intentionality. We have derived from this a claim about the logic of referring terms, namely a claim that certain terms which refer to computations are intentional-with-an-s. What is interesting about this latter claim is that we can formulate it without any talk of observers: it is applicable not just to language about human activities (such as programming and using computers) but also to, for example, the language of physics. All that is needed is to be able to distinguish between the sense and reference of terms in the language under consideration, and to examine the language to see whether there were any salient uses of opaque contexts in it.

2 The Position of this Paper

I shall be arguing for the following position. Parsing is something that happens in nature: for example, DNA processing in cells can be regarded as parsing. Now the results of parsing are things are structured objects: in order to parse something you have to use a specific set of concepts, which describe the sort of structures that the parser outputs. This requires such a set of concepts to be specified, and it will generate opaque contexts as we have seen above. So *in a sense* parsing requires a privileged "viewpoint", namely the conceptual scheme given by those concepts. But what we call a viewpoint here has none of the other conceptual apparatus of talk of observation: there is no intentionality-with-a-t in Quine's sense.

There are historical precedents for this position: as Aristotle says,

> A doctor builds a house, not *qua* doctor, but *qua* housebuilder, and turns grey, not *qua* doctor, but *qua* dark-haired. On the other hand he doctors or fails to doctor *qua* doctor. [1, Book 1 Chapter 8, 191^b1ff.]

We should note that this passage is full of opaque concepts: in fact, the *qua*$(\cdot)$ operator is one such (the doctor and the housebuilder may be the same person, but building is done *qua* housebuilder rather than *qua* doctor). So intensionality, as we have described it, seems to be relatively uncontentious.

Many areas of science rely on such a privileged set of concepts: we give examples, from logic, from statistical mechanics, from reasoning about causality and change, and from machine learning. The treatment of causal reasoning will be based on [24]. This choice of privileged concepts can be described in several ways: it can be thought of as a certain sort of intensionality, or, in logical terms, it can be described as a failure of uniform substitution. So the main body of this paper will consist of a series of examples.

3 Examples

3.1 Mereology

The mathematical theory of mereology is the formal theory of wholes and parts, and has attracted some attention as a formal foundation [20]. However, notice that mereology does not seem to be very good at talking about boundaries within things: if we can compose parts a and b to get a whole c, and if we can also make c out of d and e, then we have $a \cup b = d \cup e$. This makes it hard to see how we could talk about parsing using only mereology, because, if we parse a string (into noun phrase and verb phrase, for example) then we are saying that the boundary between noun phrase and verb phrase is the only one which counts at this level. Similarly, it is hard to talk about nesting, because the part-of relation in mereology is transitive.

These worries can be substantiated formally. Mereology on its own is equivalent to monadic second order logic, which is decidable: so if you only had mereology, you could not even parse the integers in any notation. There have been a number of philosophers of mathematics who have used mereology (Goodman, Quine, Field, Burgess, Lewis, Hellman), but they have generally augmented it with, for example, primitives for reasoning about inscriptions [6]. And these primitives for reasoning about inscription amount, I would claim, to a privileged set of concepts which must be added to the primitive language of mereology in order to do anything with it.

3.2 DNA Transcription

I shall take it (as does Endicott [4]) that the process of DNA transcription is syntactic: the syntax is given by the sequence of bases along the DNA strand, grouped into coding and non-coding DNA, and with the coding DNA grouped into triples each standing for amino acids. I shall also take it that that it has a semantics, namely the actions of protein synthesis which the DNA encodes (non-junk DNA encodes sequences of amino acids, and part at least of junk DNA seems to have a regulatory function). Precise definitions of the way in which DNA can be taken to have syntax and semantics can be found in [11], although the arguments seem to be surprisingly intricate.

We should notice that there is a certain degree of arbitrariness with DNA encoding: there is nothing in the inherent nature of base triples which means that a particular triple *must* stand for a particular amino acid. However, there are good reasons why there should be *some* standardised scheme or other, because standardised genetic machinery makes the production of viable genetic variants easier. So, once such an encoding mechanism has come about (for whatever accidental reason), there are good reasons why it should become established: this is an example of what physicists call *symmetry breaking*. Symmetry breaking is one reason why we cannot simply equate the physical with the intrinsic: symmetry breaking happens for extrinsic reasons, but there are intrinsic reasons why it should persist once established.

3.3 Statistical Mechanics

Statistical mechanics is the study of complex systems which are, at the macro scale, nondeterministic, but which are, on the micro scale, composed of large numbers of entities which interact, in many cases deterministically, in ways which can be described by the usual mechanical formalisms (either classical or quantum). So, at the micro scale, the system can be described by an evolution equation (either discrete or continuous) on a rather large phase space.

Now we are not interested in the micro scale, because we live on the macro scale, and because describing the micro scale would entail dealing with unfeasible amounts of information and would require us to work on timescales which are shorter than any changes relevant to us. We can, in typical cases, measure quantities on the macro scale (in the case of gases, for example, we measure pressure, temperature and volume) and we can predict the evolution of these macro quantities. These quantities turn out to be definable in terms of the statistical behaviour of micro-scale phenomena, and they can be genuinely predictive: as Shalizi and Moore put it,

> many physical processes … are driven by entropy increase, or by competition between maximizing two different kinds of entropy [15, p. 1]

Now this is, in a sense, paradoxical. The micro phase space can be considered to be objectively, scientifically given: however, we choose the macro variables, and these variables partition the phase space as a consequence of our choice. But we can, in many cases, formulate seemingly valid causal laws in terms of values of the macro variables: for example, the melting of an ice cube is a good example of a process driven by entropy increase, and

> These processes either happen or they dont, and observers, knowledgeable or otherwise, seem completely irrelevant. (Shalizi and Moore [15, p. 1])

These processes thus seem to be good counterexamples to subjectivist views of probability, or to views of entropy as a measure of subjective uncertainty:

> In a nutshell, the epistemic view of entropy says that an ice-cube melts when I become sufficiently ignorant of it, which is absurd. (Shalizi and Moore [15, p. 1])

3.4 The Frame Problem: Circumscription

The frame problem arises when we try to reason about change and agency in the everyday world, in which, when we bring about a change, some things change as a consequence but most things do not. This is a problem which has proven difficult to handle using the techniques of standard, extensional logic. We will describe several approaches to it: they have a common technical core, which is where the intensionality, or vocabulary-dependence, enters the picture.

Historically one of the first approaches to the frame problem uses a procedure, developed originally by McCarthy [10], called *circumscription*. We describe it using the outline in Lifschitz [7]. As is common in the literature, Lifschitz uses the notion of a fluent, namely a term whose truth value can change over time.[3] So, the procedure is as follows

1. Select appropriate fluents for the problem: for example, if we are talking about objects moving around, then we might have fluents like $\mathsf{at}(b_1, l_2, t)$, which says that block b_1 is at location l_2 at time t.
2. Secondly, “we need to describe first when a combination of values of the frame fluents is ‘consistent’, that is, [could possibly be] attained in some situation”.
3. Next we introduce actions and their postconditions: the postcondition of an action is the set of fluents which become true because it has been executed
4. There is also the general machinery of the situation calculus in particular the generic law of inertia

   $$\neg\mathsf{noninertial}(f,t) \quad \rightarrow \quad [f(t+1) \leftrightarrow f(t+1)]$$

   (this is a somewhat simplified version of [7, 5.14, p. 333]; here $\mathsf{noninertial}$ is a predicate which says that a fluent has changed its truth value at a particular time).
5. Finally we compute the resulting solution by circumscription: we circumscribe – that is, minimise – the extent of the $\mathsf{noninertial}$ predicate while holding fixed the postconditions of the actions and obeying the constraints.

3.4.1 What are Fluents?

First a question: what are fluents? On the one hand, they *look* very propositional – they are such things as $\mathsf{at}(b_1, l_2, t)$ – and, given a fluent $f(\ldots, t)$ and a time t, we can obtain a truth-value. So they might, conceivably, be families of propositions parametrised by times (and, indeed, they are called “propositional fluents”: [7, p. 328]).

However, the propositional appearance is deceptive: they are, as Lifschitz remarks, “terms and not formulae” [7, p. 328]. We cannot, for example, apply truth-functional connectives to them: in particular, closing them under disjunctions is a good way to make the circumscription mechanism break down.[4]

So fluents have two sides: extensional and intensional. Extensionally, they can be regarded as assignments of truth values to situations, and, regarded in this way, there would be no reason why we should not form arbitrary truth-functional combinations of fluents. Intensionally, however, fluents are terms: we cannot necessarily form arbitrary truth-functional combinations of them.

3.4.2 The Failure of Uniform Substitution

There is another way to look at this intensional side. Fluents are typically taken to be *literals*, and the choice of literals then comes down to the choice of suitable primitives for our language. However, if we do this as a matter of policy, then we cannot choose different primitives: we can have different, but logically equivalent, languages, which yield, on application of the circumscription procedure, different results, precisely because the fluents are different in each case. Thus, a logic formulated using circumscription is not closed under uniform substitution. This is something which has been known for some time, and not usually regarded as a significant problem – see [8]: one could think of this as a cause for concern, since it makes our deductions about the world dependent on the primitives that we use to express them in. However, as we argue, we generally have to use a specific vocabulary for causal reasoning, so this should not be any surprise.

And it turns out that the choice of primitives has a considerable influence on the outcome of circumscription: we can show that, by choosing the primitives appropriately, we can (in the propositional case at least) make the outcome of circumscription be *anything we want* consistent with the axioms and the action postconditions [21].

---

[3] Many authors in fact use fluents whose value depends on situations, where situations are nodes in the tree of all possible action sequences: we will use the temporally indexed version, because it shows all of the phenomena that we are interested in, it is much easier to align with philosophical work on the grue paradox, and it is notationally much simpler: we have a great deal of notation already, and introducing a tree of situations in addition would be needlessly complicated.

[4] We may, of course, define what Shanahan calls “compound fluents” [16, pp. 115f.], and we can define recursive conditions for $f(\ldots, s)$ to be true when f is such a fluent. However, these fluents play no role in the minimisation procedure; and it is this role in the minimisation procedure that we are concerned with.

Clearly a great deal is being smuggled in under the choice of primitives, and it would be good to have some more mathematical insight into what the choice amounted to. We will do this in the next section.

3.4.3 The Frame Problem: Explanation

McCain and Turner [9, 19] found out how to reformulate McCarthy's circumscription procedure in a mathematically more perspicuous way. We give ourselves a language $\mathcal{L}$, together with a set Ξ of clauses, which I shall write in the form

$$\phi \rhd \psi$$

Here ϕ and ψ are propositions, time-varying in the case of the frame problem. McCain and Turner have a model-theoretic (and rather complex) definition of inference with these clauses: it is equivalent to the following.

First we define a modal operator $\Box_\Xi$, depending on Ξ: it will be the strongest **K** modal operator on $\mathcal{L}$ for which all of the entailments

$$\phi \vdash \Box\psi \quad \text{for all } \phi \rhd \psi$$

are theorems. We have given a proof theory for this operator which satisfies cut elimination [22, 23]; the left and right rules for the modal operators are Table 1. We then add to our theory the set of axioms $p \leftrightarrow \Box p$, for all propositions p.

$$\frac{\{\Gamma, \phi_1, \ldots, \phi_k \vdash \Delta\}_{\psi_1,\ldots,\psi_k \vdash P}}{\Gamma, \Box_\Xi P \vdash \Delta}\ \Box\mathrm{L}$$

$$\frac{\Gamma \vdash \phi_1 \wedge \ldots \wedge \phi_k, \Delta \qquad \psi_1, \ldots, \psi_k \vdash P}{\Gamma \vdash \Box_\Xi P, \Delta}\ \Box\mathrm{R}$$

Here we have $\phi \rhd \psi$ for all relevant ϕ, ψ; in $\Box$L the indexing is over a set of ψ_i with $\psi_1, \ldots, \psi_k \vdash P$ where that set is downwards closed in the set of all such under containment.

Table 1. Proof Rules for the McCain-Turner Modals

Remark 1. Notice that cut elimination is important here: *prima facie* the left rule has an infinite number of antecedents, but (if the clauses are well enough behaved) an argument analogous to that for the subformula property shows that a finite number of them suffice.

Remark 2. Although the original application to the frame problem only requires clauses with non-modal members, the proof of cut elimination applies to the case where we define well-founded sequences of modalities by means of clauses involving modalities earlier in the sequence. This will be important for our results on the semantics of classical proofs.

We can regard this logic as describing a certain sort of explanation: the clauses in Ξ can be regarded as basic explanations, and an entailment

$$p \vdash \Box q$$

can be read as "p explains q". And, when we add the axioms $p \leftrightarrow \Box p$, we are endeavouring to construct an *explanatorily closed* theory: that is, we have a language which contains not just propositions, but explanations of those propositions, and we want a description of the world in which every fact has an explanation and in which every explanation guarantees the truth of the explained fact.

We will, throughout this paper, use a terminology of explanation in describing the McCain-Turner theory, rather than, as they do, using a terminology of causality: in particular, sets of clauses will be called *explanatory*, rather than causal, theories.

So much for the general theory. We apply it to the frame problem by considering sets of clauses like those in Table 3 (see [9]). This is something of an improvement over circumscription, because the action takes place in a monotonic modal theory: the only part of the machinery which is nonmonotonic turns out to be that, if you add more clauses to Ξ, you can end up with fewer valid inferences than you started with (because of the left rule for $\Box_\Xi$, or the right rule for $\Diamond_\Xi$). And generally the logic is better behaved: we can prove cut elimination, and cut elimination leads to the possibility of proof search.

There will be two types of propositions here, fluents and actions: We write fluents and actions with a temporal index, but this is not part of the formal syntax(we cannot quantify over the indices, for example).

First, we give initial conditions at $t = 0$: that is, for fluents or negations of fluents p_0 at time 0, we have

$$\rhd p_0 \tag{6}$$

next, we say that, if an action occurs at t, it explains its postconditions at $t + 1$: so, if p is the postcondition of α, we have

$$\alpha_t \rhd p_{t+1} \tag{7}$$

for all t. Finally, we give the frame axiom: for all fluents ϕ, we have

$$\phi_t \wedge \phi_{t+1} \rhd \phi_{t+1} \tag{8}$$

or, in other words, if a fluent is true at t, and its truth value is unaltered between t and $t + 1$, that is sufficient explanation for its truth at $t + 1$ (in other words, for fluents persistence is self-explanatory).

Table 3. McCain and Turner's "Causal" Rules

However, there is still one part of this construction which is not obviously invariant under uniform substitution: that is, the choice of fluents. Arguably we have still made progress, because the fluents figure in the clauses defining the basic explanations, and we use the basic explanations as a presentation of the modal operator, which defines our notion of (non-basic) explanation. And, in general, there will be different choices of fluents which give rise to the same modal operator, and so the McCain-Turner formulation is not as brutely nonextensional as the standard formulation of circumscription. However, as the examples in [21] show, there are choices of fluents which do change the outcome of circumscription, and so, if we reformulate this circumscription using McCain-Turner's formalism, we get choices of clauses which do change the modal operator (since the modal operator determines the predictions in the McCain-Turner formalism). So, clearly, there is something here still to be investigated.

$\dfrac{}{\Gamma, \bot \vdash \Delta}\ \bot\,\mathrm{L}$	$\dfrac{}{\Gamma \vdash \top, \Delta}\ \top\mathrm{R}$
$\dfrac{\Gamma, P, P' \vdash \Delta}{\Gamma, P \wedge P' \vdash \Delta}\ \wedge\mathrm{L}$	$\dfrac{\Gamma \vdash Q, \Delta \qquad \Gamma' \vdash Q', \Delta'}{\Gamma, \Gamma' \vdash Q \wedge Q', \Delta, \Delta'}\ \wedge\mathrm{R}$
$\dfrac{\Gamma, P \vdash \Delta \qquad \Gamma', P' \vdash \Delta'}{\Gamma, \Gamma', P \vee P' \vdash \Delta, \Delta'}\ \vee\mathrm{L}$	$\dfrac{\Gamma \vdash Q, Q', \Delta}{\Gamma \vdash Q \vee Q', \Delta}\ \vee\mathrm{R}$
$\dfrac{\Gamma \vdash Q, \Delta}{\Gamma, \neg Q \vdash \Delta}\ \neg\mathrm{L}$	$\dfrac{\Gamma, P \vdash \Delta}{\Gamma \vdash \neg P, \Delta}\ \neg\mathrm{R}$
$\dfrac{\Gamma \vdash \Delta}{\Gamma, P \vdash \Delta}\ \mathrm{WL}$	$\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash Q, \Delta}\ \mathrm{WR}$
$\dfrac{\Gamma, P, P \vdash \Delta}{\Gamma, P \vdash \Delta}\ \mathrm{CL}$	$\dfrac{\Gamma \vdash Q, Q, \Delta}{\Gamma \vdash Q, \Delta}\ \mathrm{CR}$
In the Axiom rule, A is non-modal.	

Table 2. Classical Rules for the McCain-Turner Modal System

3.4.4 The Frame Problem: The Grue Paradox

We have, then, reduced the frame problem to a single question: that of finding fluents, i.e. time-dependent primitives for which persistence is explanatory. Now if we transpose this problem into the terms of the problem of prediction, we get the following question: for which temporally parametrised propositions can observation license prediction? In this formulation, the problem has been known in the philosophical community, and is officially referred to as Nelson Goodman's "New Riddle of Induction", and more colloquially as the "grue paradox". [18, 17] [5, Ch. III] [13]

Abstractly, the problem can be described as follows. Suppose that we have a language with two temporally indexed propositions, ϕ and ψ, and that ϕ and ψ are contradictory (i.e. that, for all t, $\phi_t \wedge \psi_t \vdash \bot$). Suppose also that we can successfully predict ϕ and ψ by observation: that is, that the observation of ϕ_t warrants the prediction of ϕ_{t+1}, and similarly for ψ. Then, for some t_0, define[5]

$$\tilde{\phi} = \begin{cases} \phi & (t < t_0) \\ \psi & (t \geq t_0) \end{cases}$$

$$\tilde{\psi} = \begin{cases} \psi & (t < t_0) \\ \phi & (t \geq t_0) \end{cases}$$

We note the following:

1. If ϕ and ψ warrant prediction, then $\tilde{\phi}$ and $\tilde{\psi}$ cannot, because the two pairs make contradictory predictions
2. The relation between the $\langle \phi, \psi \rangle$ pair and the $\langle \tilde{\phi}, \tilde{\psi} \rangle$ pair is entirely symmetric: the former can be defined from the latter by means of definitions with entirely the same form as the definitions above
3. We can, then, regard the language in which these definitions are formulated in two ways: one in which $\langle \phi, \psi \rangle$ are primitive and in which $\langle \tilde{\phi}, \tilde{\psi} \rangle$ are defined, and *vice versa*. These two formulations are indistinguishable from the point of view of logical form
4. Attempts to say that $\langle \tilde{\phi}, \tilde{\psi} \rangle$ are somehow "artificial" because their definitions involve an arbitrary time are question-begging: their definitions only involve an arbitrary time *in a language in which the rival concepts are primitive*
5. Similarly, attempts to say that our mental concepts, or the recognitional capacities exercised by our senses, legitimate $\langle \phi, \psi \rangle$ rather that $\langle \tilde{\phi}, \tilde{\psi} \rangle$ are likewise question-begging: how do we know that the concept we now deploy is ϕ or $\tilde{\phi}$?.[6] For this reasons, attempts such as this are liable to be received with a certain amount of scepticism by anyone familiar with the literature on Wittgenstein's Private Language Argument

The grue paradox, then, is non-trivial. The more or less received philosophical position is that, in order to predict change, one needs to have a set of primitives which, in the absence of causes, will persist unchanged: such primitives are called *projectibles*. Sets of projectibles cannot be chosen on purely logical grounds, but must, in some way, reflect the causal structure, or ontology, of the world; that is what the grue paradox shows.

We should note that the grue paradox, traditionally stated, is about the licensing of inference by projectibles, whereas, as we have argued, the McCain-Turner formalism is about the licensing of *explanation* by projectibles. The basic pattern is that, if P has the same truth value at 0 and 1, and P is true at 0, then we have an explanation of P being true at 1. Explanation seems in any case to be more delicate: for example, logically valid inference seems to be closed under disjunction, whereas explanatory projectibles are clearly not.

We have, so far, a language consisting of temporally indexed propositions, of which there are two sorts, actions and fluents. We can define new primitives in a three-sorted language: there will be

---

[5] Traditionally, $\tilde{\phi}$ is known as *grue*, and $\tilde{\psi}$ as *bleen*.

[6] [2] has a system in which he allows the agent *to know which concepts they are deploying*, and this – unsurprisingly – resolves the grue paradox.

times $(t_0, t_1, \ldots)$ and trajectories, which will intuitively speaking be n-tuples of fluent literals $\lambda = (f_0, f_1, \ldots)$ (supposed finite for simplicity). We can recover a temporally indexed proposition from a trajectory and a time: $\lambda \wedge t_i$ will be some temporally indexed proposition P_i. Similarly, we can define times: t_0 will be $f_0 \vee f'_0 \vee f''_0 \ldots$, where we make a disjunction of all of the fluent literals indexed with 0. And we can make a trajectory from a series of fluent literals: the trajectory will be of the form $(t_0 \rightarrow f_0) \wedge (t_1 \rightarrow f_1) \wedge \ldots$. So the language of trajectories plus times is equivalent to the simple propositional language that we started with. Using this equivalence, we can define operations on trajectories, using merely the logical operations on propositions: we can modify a single temporal value of a trajectory, or, given two trajectories and a time, we can graft the two trajectories together: λ_1 before t, λ_2 at and after t. We can, thus, start with a set of *primitive* trajectories and construct the others from them: in this way, the two alternative futures for the frame problem correspond to different choices of primitive trajectories.

We can now rewrite the McCain-Turner set of rules in terms of trajectories. The initial conditions are no trouble: they are all of the form $\triangleright(t_0 \rightarrow \lambda)$, for suitable λ. Similarly, the action postconditions will all be of the form $(t \rightarrow a) \triangleright (t + 1 \rightarrow p)$, for suitable t, a and p. The persistence axioms, however, will all be of the form

$$(t \rightarrow \lambda) \wedge (t + 1 \rightarrow \lambda) \triangleright (t + 1 \rightarrow \lambda) \qquad (9)$$

and, simply by varying the primitive trajectories, we get one outcome or the other for the grue paradox. So we have shown:

The McCain-Turner system reduces the frame problem to the grue paradox.

In this context, the McCain-Turner system seems to be rather important: it has swept away a large amount of what appear to be *ad hoc* constructions and focussed on a philosophically significant issue. We shall now concentrate on that issue.

3.4.5 Parsing

We can simply note here that parsing can be fitted into the McCain-Turner framework: if we are parsing English, for example, we can write rules which explain the syntax of a sentence by a concatenation of noun phrase and verb phrase, explain a noun phrase as a concatenation of determiner and noun, and so on. We get, in this way, an encoding of the Lambek calculus into this modal logic.

3.5 Machine Learning

We have seen that logical reasoning about change is intensional, in that the logical constructions involved are not stable under uniform substitution: similarly, explanation is intensional in that some terms will figure in good explanations (and particularly explanations of change) whereas other terms, which might well be logical combinations of the good ones, will not be similarly explanatory.

Exactly similar issues raise themselves for machine learning. Suppose that we have a classification problem: that is, we have a population of individuals, which we want to group together on the basis of their *features*, that is, the values of certain specified attributes. So we can take a sample of the population, measure the values of the specified attributes, and cluster them using these features using some clustering algorithm or other: the output of the clustering algorithm will give us a classifier for the population. Again we find that the output of algorithms like these is sensitive to the precise feature set chosen: in fact, the precise choice of suitable feature sets – or the development of theoretical frameworks which show how this choice should be performed – is one of the main tasks in the practice of machine learning.

4 Conclusion

So our position on Searle's views on syntax has two components. The first is that there is, indeed, a certain intensionality in parsing: it is, however, not correctly described as observer dependence. This sort of intensionality is quite pervasive in science, and consideration of non-syntactic examples gives one a better perspective on many of the issues (objectivity, naturalism, and observer-relativity) than is given by looking at parsing on its own. We have shown, by a series of examples, that there is a pervasive intensionality underlying a great deal of causal explanation: this intensionality is, we should argue, the basic intuition behind Searle's position on syntax, rather than anything to do with observer-relativity. To go beyond this, and, from intensionality, to conclude observer-relativity, is making a large and, I would argue, unwarrantable leap.

However, there are also unexplained phenomena here. The pervasiveness, in physics and other sciences, of intensionality is somewhat counterintuitive, and needs better explanation than we have at present. We should note that the first section of [15] is entitled "What's strange about macrostates, or is it just me?": the counterintuitiveness of these phenomena is felt, it seems, by even professionals in these fields.

REFERENCES

[1] Aristotle, *Physics*, Princeton, 1984. In *The Complete Works of Aristotle*, vol. I, ed. J. Barnes.

[2] John Bell, 'A pragmatic theory of induction', in *Proceedings of ECAI 2002*, (2002).

[3] Selmer Bringsjord and Ron Noel, 'Real robots and the missing thought-experiment in the chinese room dialectic', In Preston and Bishop [12], 144–166.

[4] Ronald P. Endicott, 'Searle, syntax, and observer-relativity', *Canadian Journal of Philosophy*, **26**(1), 101–122, (1996).

[5] Nelson Goodman, *Fact, Fiction and Forecast*, Harvard University Press, fourth edn., 1983.

[6] Geoffrey Hellman, 'Mereology in the philosophy of mathematics', in *Handbook of Mereology*, eds., Hans Burkhardt, Johanna Seibt, and Guido Imaguire, Philosophia Verlag, (2009).

[7] Vladimir Lifschitz, 'Circumscription', in *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, 297–352, Clarendon, (1994). Volume Coordinator D. Nute.

[8] David Makinson, 'Ways of doing logic: What was different about AGM 1985?', *Journal of Logic and Computation*, **13**(1), 3–13, (2003).

[9] Norman McCain and Hudson Turner, 'Causal theories of action and change', in *Proceedings of AAAI97*, pp. 460–465, Providence, RI, (1997). AAAI.

[10] John McCarthy, 'Circumscription – a form of non-monotonic reasoning', *Artificial Intelligence*, **13**, 27–39, (1980).

[11] Ruth Garrett Millikan, *Language, Thought and Other Biological Categories*, MIT Press, 1984.

[12] *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, eds., John Preston and Mark Bishop, Oxford University Press, 2002.

[13] Hilary Putnam, 'Foreward', In *Fact, Fiction and Forecast* [5], vii–xvi.

[14] John R. Searle, *The Rediscovery of the Mind*, MIT Press, 1994.

[15] Cosma Rohilla Shalizi and Christopher Moore, 'What is a macrostate? subjective observations and objective dynamics'. `http://arXiv.org/abs/cond-mat/0303625v1`, 2003.

[16] Murray Shanahan, *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, MIT Press, 1997.

[17] Douglas Stalker, 'Grue! the new riddle of induction', *Canadian Philosophical Reviews*, **15**(3), 211–213, (June 1995).

[18] *Grue! The New Riddle of Induction*, ed., Douglas F. Stalker, Open Court, Peru, 1994.

[19] Hudson Turner, 'A theory of universal causation', *Artificial Intelligence*, **113**, 87–123, (1999).

[20] Achille Varzi, 'Mereology', in *The Stanford Encyclopedia of Philosophy*, ed., Edward N. Zalta, winter 2012 edn., (2012).

[21] G. Graham White, 'Intensionality and circumscription', in *Proceedings of the 9th International Conference on Non-Monotonic Reasoning (NMR2002)*, eds., Salem Benferhat and Enrico Giunchiglia, pp. 372–379, (2002). Toulouse, April 2002.

[22] G. Graham White, 'Interpolation for McCain-Turner causal theories', in *Proceedings of NRAC'03*, (2003).

[23] G. Graham White, 'Causality, modality and explanation', *Notre Dame Journal of Formal Logic*, **49**(3), 313–343, (2008).

[24] G. Graham White, 'The frame problem and the semantics of classical proofs', in *Alan Turing Centenary: The Alan Turing Centenary Conference*, ed., A. Voronkov, pp. 415–429, (2012).