

# Reading the Past in the Present

Nick Huggett, University of Illinois at Chicago  
huggett@uic.edu

July 25, 2014

## 1 The Knowledge Asymmetry in *Time and Chance*

### 1.1 Memories and Macroscopic Traces

Why is our knowledge of the past so much more ‘expansive’ (to pick a suitably vague term) than our knowledge of the future? And intimately related, how can we capture the difference(s): i.e., in what sense is knowledge of the past more ‘expansive’? As a first stab, one might be convinced by the first four chapters of David Albert’s *Time and Chance* (2000; henceforth T&C – all quotations are from this work unless otherwise stated) that the ‘Newtonian statistical mechanical contraption for making inferences about the world’ (96) captures everything that can be inferred by statistical mechanics at any time. Indeed, in this paper I will assume that it does. One might then reasonably wonder whether it would even be possible to know anything that didn’t follow from the contraption – so doing would apparently take more knowledge of the past than the past hypothesis (PH), or more knowledge of the present state than its current macrocondition, or a more informative probability distribution, and how could one obtain any of those? So one might suggest that:

*A proposition can only be known if it is made likely by taking a uniform probability over all states compatible with the current macrocondition of the universe, and with its initial macrocondition, given the laws of mechanics.*

Then, since the initial state is one of considerable ‘order’, given by the PH<sup>1</sup>, many more interesting propositions about the past than about the future will satisfy the necessary condition .

Such a necessary condition can’t be the end of the story: it is asymmetric, permitting more knowledge of the past than future, but it doesn’t make clear how we come to have any such knowledge. For instance, it would be manifestly crazy to suggest that the condition describes how we reason about other times, because clearly we have knowledge of other times without knowing anything much about the nature of the early universe, and so presumably has humanity since before the earliest creation myths. Our ability to reconstruct the past is not evidence of our knowledge of the PH! But if it were a necessary condition (it’s not) then it could be a great help: having a decent statement of the asymmetry surely aids explaining it! And it relies on the PH, so any satisfactory account of the knowledge asymmetry must explain why; likely such an account will invoke the PH.

---

<sup>1</sup>Two points: first, the PH is not merely that the universe started with ‘low entropy’ (a claim that some say is ‘not even false’) but that it started in some, to be specified by the full PH, macrostate. Second, I will make the usual assumption of an even probability function over microstates throughout this essay, often without further comment for brevity.

And in fact, the treatment of the knowledge asymmetry in T&C (Chapter Six), does proceed in much this way. It places giving a clear description of the asymmetry forefront in the discussion, as a prerequisite to an explanation, and so presents and defends a necessary condition along these lines. However, the condition is significantly different in an important way that is not made terribly clear in T&C: my main goal in the first half of this paper is to explain and clarify the difference and say why it is so important. Also different is T&C's argument for the condition, as I'll explain towards the end of this half; in the second half, with a clear account of the approach before us, I will argue that it doesn't succeed. While I have no positive account to offer in its place, I hope at least to make some progress by illuminating some of the constraints on any proposed understanding of the knowledge asymmetry. In particular, the focus of the paper is on getting straight(er) about the physics of *memory*.

So, suppose the 'ice pachinko' of T&C (83) is run with a single ice cube in a sealed room with a human observer watching, but making no record of events. The cube falls through the device and randomly ends up in one of the beakers, the leftmost one say. Eventually, after the ice cube has first melted and then evaporated, there will be no trace of which beaker it fell into remaining in the macrostate of the universe: the water molecules will be randomly distributed in the air of the room, whatever beaker the ice cube fell into. The macrocondition of the universe will be the same whatever happened, and the initial macrostate of the universe surely won't determine the outcome (certainly it isn't supposed to do that kind of work in T&C), so the contraption should assign equal conditional probabilities for the ice having falling left or right – that it fell left is not knowable according to our necessary condition.

That's not to say that the *microstate* doesn't contain the necessary information: if the evolution is (backwards) deterministic, then the earlier state can be recovered from the microstate. But the whole game here is to work with macrofacts – idealizing, we take those to be what's knowable. What about the fact that we can discover aspects of the microstate: perhaps there are insensible mineral traces in the beaker that could be detected by chemical analysis? Such discoveries amount to magnifying micro-differences to produce macro-differences: paradigmatically, by focussing a microscope on something otherwise too small to see. But the proposed formula for the knowable only takes into account the *present* macrocondition; that some aspect of the microstate could or will be magnified into the macrostate, at a later time is irrelevant. The necessary condition says that the outcome is knowable now only if it can be inferred using the macrostate now, and it is beside the point to note that it could be satisfied if the macrostate were different, or even that it will be satisfied by some future macrostate. It isn't satisfied at the time in question.

But of course, that the ice cube fell in the leftmost beaker is perfectly knowable at that time – it is already known by the observer! So the condition is not necessary, and it cannot help explain the knowledge asymmetry. And neither does this example seem especially *recherché*. Macrorecords are being erased all the time, and the observer is relying on no strange procedures for obtain knowledge – she's just watching!

## 1.2 The 'Presently Surveyable Condition'

As I said, the account just refuted is not that offered in T&C: instead of conditioning on the current 'macrocondition', T&C conditions on (what it terms elsewhere) the 'presently surveyable condition' of the world:

“... everything we know of the past and present and future history of the world can be deduced ... from the following four elements: what we know of the world's present

macrocondition – and of our own brains, perhaps; the standard microstatistical rule; the dynamical equations of motion; the past hypothesis.” (119)

In fact, the ‘Newtonian contraction’ uses exactly these elements, not those in the condition I gave; we can say for short that we can know something of the past or future if it is given a high probability by the contraction. (Apologies for the earlier misdirection, which I found useful for setting things up.)

So the difference is that we are also to condition somehow on the physical state of our brains. Before I elaborate this crucial idea, I do want to emphasize that it is rather easy to overlook in T&C. The passage quoted downplays its relevance; and it first appears with almost no comment or explanation (96). The fullest discussion appears in a footnote: the surveyable condition includes ‘whatever (perhaps [microscopic]<sup>2</sup>) features of the present condition of the *brain* of the observer in question may be accessible to her by means of direct introspection.’ (114) But, as we’re about to see, if it isn’t obvious already, conditioning on the brain state in some way is essential in this project, in order to address mnemonic knowledge; T&C is entirely right to require such conditioning.

The idea is that the pachinko observer’s knowledge satisfies this new condition: that her memories entail something about the microstate of her brain that, with the PH (and the rest of the macrocondition), make it likely that the ice fell in the leftmost beaker. The previous example should make clear why some such conditioning is absolutely necessary if we are to include memories of the past in our treatment of the knowledge asymmetry – and surely one of the things we want to know most is why we remember the past but not the future! ‘Direct introspection’ may be intended to be more expansive, but in the case in hand all we need add to the surveyable condition are our memories – or rather, since we will have to worry about their veracity, our putative memories.<sup>3</sup> So we include whatever follows for the physical state of the world (including its microstate) from the fact that the pachinko observer has a (putative) memory of the ice cube falling into the leftmost beaker.

By ‘putative’ I mean that all that’s relevant is the fact that the observer’s brain is in a state *compatible* with her truly remembering the beaker; of course just being in that state is no guarantee that the memory is veridical. Especially we must bear in mind that the background to this discussion is the reversibility objection: if we accept that future entropy increase is overwhelmingly likely, conditional only on the present surveyable state and the laws, then the symmetry of the laws implies that the conditional probability for past entropy *decrease* is equally likely. So in this discussion we need to entertain the possibility that instead of being veridical, putative memories are the result of extremely improbable fluctuations in the microstate of the world. (Let’s bracket other sources of false memories momentarily.) Of course, if one conditions on *veridical* memories then, by logic rather than physics, it follows that the remembered events really happened. For example, that the pachinko observer veridically remembers that the ice cube fell in the leftmost beaker, entails that it did. But the reversibility objection shows that there can be nothing about the present microstate of a brain with some memory that entails its veracity – at very least, the stimuli that eventually produced the memory are most likely the product of random fluctuations, not the recalled event. So if we want to play the usual game (that of T&C) of assuming about the present only that which

<sup>2</sup>The text reads ‘macroscopic’, which is obviously a typo; compare with (96).

<sup>3</sup>Here’s another reservation about the treatment of the brain state in T&C. Talk of ‘direct introspection’ carries a suggestion of armchair psychology: it sounds as if what counts is ‘second order’ knowledge of what our putative memories are. But the discussion should make clear that what actually matters is just the memories themselves – not whether we also know we remember in some sense. I’m pretty sure David is aware of this, but for a while I was confused by the language.

could (in a very broad sense) be learned by inspection of the universe now, we can't condition on the truth of memories, only their existence. Put another way, to know that a memory is veridical is to know something about the state of the past that doesn't follow from the present state and the laws, and including such information undermines T&C's whole project of understanding time asymmetries in terms of statistical mechanics. (Thus a note on terminology: the unmodified term 'memory' will now mean 'putative memory'.)

So the necessary condition for knowledge is just as in the previous section, except that we further condition on whatever follows for the state (micro or macro) of the world from the observer's putative memory of the ice cube falling in the leftmost beaker. The idea is that there are only two explanations for the state of the brain on the table, random fluctuation or faithful recording process, and the PH makes the latter far more likely. (As we'll see, there is a significant problem here: what exactly does a memory entail for the state of the brain? And does whatever it is actually imply that these are the only possibilities?)

Assuming (against some critics) that the Newtonian contraption generally works as advertised, it still might seem implausible that just conditioning on an appropriate brain state could be enough to significantly raise the probability that the ice cube fell in a particular beaker. After all, there are lots of ways in which such belief formation might go wrong – optical illusions, tricky magicians, evil scientists, wishful thinking, forgetfulness and so on. But it's worth noting that non-veridical memories formed by such processes may not make their subjects likely, if such processes leave other traces in the presently surveyable condition. To give a fanciful example, consider the group of GIs 'brainwashed' to hold false memories about the heroic actions of Raymond Shaw, in the *Manchurian Candidate*. Applying the laws to those memories (consistent with the PH), we assume for now, makes the true past unlikely; conditioning on the states of their brains (and the PH) makes it likely that Shaw was a hero. But whether or not he really was makes a difference to the present surveyable condition – beyond those memories – in the memories of the Soviet observers, in records kept by the North Koreans, in messages sent to conspirators in the US, and so on. If those traces are present, and if the contraption is working, it ought to make the beliefs unlikely to be true, after all. And more generally, it's easy to imagine cases, like the pachinko observer, in which a veridical memory is the only effect of an event on the presently surveyable condition: watching how a wave breaks, or the milk spills, or where the first raindrop lands, and so on. But it is much harder to imagine the formation of a non-veridical memory (given the PH) leaving no traces: conspiracies require work, and work leaves traces. So it's plausible that the Newtonian contraption is generally good at assigning low probability to non-veridical memories – and in fact, it should be fairly obvious that the kinds of remarks just made apply, not only to memories, but to any kind of traces of the past.<sup>4</sup> That said, such examples will not concern us further as we proceed.

### 1.3 Records and the PH

Now that we understand the necessary condition more clearly, we can turn to T&C's argument for it. The strategy is to give an idealized account of the methods by which knowledge of other times can be had, thus obtaining (asymmetric) upper bounds on what can be known of past and future. One method for obtaining such knowledge involves inferences from the laws and surveyable condition

---

<sup>4</sup>And what about a case in which the Newtonian contraption makes likely the subject of a false memory, because traces of the deception have been erased? Would that be a counter-example to the proposals of T&C? No, because ultimately the game is to balance our beliefs and theory so that our beliefs turn out to be likely true according to the theory – and any program like that, since it is probabilistic, will end up making some false things true.

(and standard measure) alone; that generally gets the future right, but of course is radically wrong about the past. (T&C has a nice discussion of why this method doesn't get the future quite right either, and how to fix the problem [119-22] but we don't need to take account of that here.)

The second method is that of record reading: comparing the current state of a system with its state at another time (the 'ready state'), to infer (using the dynamical laws and uniform probability) the occurrence of an interaction in between the two times – at an earlier time than the present, if the ready state occurred even earlier. To literally, to explicitly make such an inference requires knowledge, not only of the current state, but of the ready state as well. Whether there is some other non-literal, implicit procedure for obtaining knowledge from a record is the subject of the second half of this paper. The argument of T&C, however, does invoke a procedure of literal inference (without prejudice concerning the existence of other methods):

And the puzzle is about how it is that we ever manage to *come* by [information about the ready state of a record bearing system]. It can't be by means of retrodiction/prediction . . . . It must be because we have a *record* of that other condition! But how is it that the ready condition of this *second* device is established? And so on (obviously) ad infinitum. There must be . . . something we can be in a position to *assume* about some other time . . . the mother (as it were) of all ready conditions. And this mother must be *prior in time* to everything of which we can potentially ever *have* a record, which is to say it can be nothing other than the initial macrocondition of the universe as a whole.

And so it turns out that *precisely* the thing that makes it the case that the second law of thermodynamics is (statistically) true . . . is *also* the thing that makes it the case that we can have epistemic access to the past which is not of a predictive/retrodictive sort [i.e., the PH]. (T&C 117-8)

That is, T&C gives a regress argument to conclude that knowledge of the literal record reading kind requires our assuming the PH. If so, the basis of knowledge of other times is the presently surveyable condition, the PH and the laws: what we can know of other times is given by the consequences of the Newtonian contraption, as T&C's necessary condition says.

Ok . . . but then, since we do have non-retrodictive knowledge of the past, knowledge presumably garnered from records, and could in almost complete ignorance of the PH, the regress just shows that we aren't just literal record readers. (Or perhaps we subconsciously know the PH, and are literal record readers – so we need psychology not cosmology to discover the PH!)

I hope David will forgive me having a little fun here. Of course that's not what's intended – as he's pointed out repeatedly, you'd have to think him crazy to take the argument that way! Instead we have to bear in mind that we are considering an idealized procedure, capable of capturing anything knowable, not an account of how *we* know.<sup>5</sup> That said, the passage perhaps invites misinterpretation by speaking in the first person; perhaps it would be clearer if it asked how 'an idealized system' that employed the method of literal record reading could come by information about ready states.

A procedure that works that way is logically consistent, but there are a couple of important points, which will feature prominently in the second half of the paper. First, since the regress argument shows that literal record reading is not only the procedure by which we obtain knowledge

---

<sup>5</sup>It's true that T&C doesn't say this explicitly, but (a) the argument to sanity, and (b) the footnote on (116) describing retrodiction in such terms, shows that is what's intended.

of the past, how do we? Until we have an answer to that question, the argument makes knowledge of the past a bit of a paradox – the resolution is not difficult, but seeing clearly how we do use records will advance the discussion. Second, taking the argument as I have just suggested reveals a logical gap: since there must be other procedures, why think that literal record reading is maximal? In other words, why think that T&C’s necessary condition holds?

Now, approaches to the knowledge asymmetry that appeal to the second law are familiar, and T&C grounds the second law on the PH, so one might suspect that we have here an appeal to entropy in understanding the knowledge asymmetry. (Especially since T&C contains another, similar regress argument – the one involving the ice pachinko – to argue that the PH grounds the second law.) But the argument is supposed to be a novel one, breaking from views which assume that records are inevitably entropy increasing. First, such an assumption is not made in the argument. But more importantly, as T&C points out, there is no way to connect its conception of a record up with the second law: whether the final state is of higher or lower entropy than the ready state, or before or after it, knowledge of the two states will generally allow one to infer something about the intervening time. (Even knowledge that both states are equilibrium for a system allows one to infer that nothing happened to the system in between!)

So things are set up for the final sections now. First, we saw that the kind of account of the knowledge asymmetry developed in T&C must include our (putative) memories in the ‘currently surveyable state’ – for there can readily be things we remember about the past that don’t leave any traces in the present macrostate. Now we have seen how T&C characterizes knowledge of other times that is not of the prediction/retrodiction sort – not in terms of the actual processes by which we know the past, rather by reference to a hypothetical process of literal record reading. But how do we actually know things about the past? In particular, how should we understand *memory* in this framework? It seems to me that until we have thought about the actual processes it is not obvious *a priori* that being knowable through hypothetical record reading is a necessary condition.

## 2 Memories Are Made of This

### 2.1 An Information Gathering and Utilizing System

One thing that may strike you in thinking about record readings is that when we remember, there’s no consciousness of record reading going on at all: you don’t generally note that you have a memory, and *infer* from that that something occurred – generally, part of a memory just is the belief that its object occurred.<sup>6</sup> So if we can get a handle on such mnemonic knowledge, perhaps that will give us an ‘inference method’ quite different from literal record reading. Anyway, memories are both central to the knowledge asymmetry, and in some way surely critical to our having knowledge of the past. So they are a good topic to investigate. We’ll approach things with a simplified model of the physical basis of our mnemonic knowledge, both as a test case for T&C in its own right, but also under the assumption that it provides a reasonable model of our mnemonic knowledge. I have in mind some version of Gell-Mann’s (1994) “information gathering and utilizing system”, or *IGUS*.

In general terms, such a device has sensors capable of responding to its environment and devices capable of affecting its environment: inputs and outputs broadly construed. The system takes inputs from its environment (visual inputs, for example), processes the received information and

---

<sup>6</sup>Of course that’s not always true: it makes perfect sense to doubt a putative memory, and to think about what might have caused it instead of its apparent object. But we’ll stick to the straight-forward cases here.

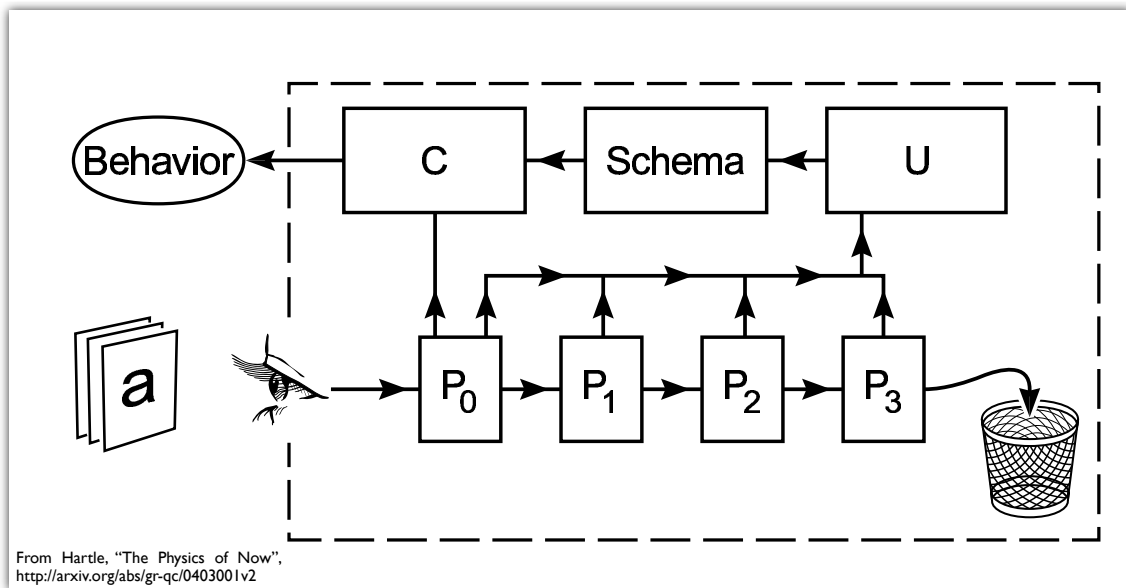


Figure 1: Hartle's 'Information Gathering and Utilizing System'. At each time: inputs are shuffled down the registers  $P_0 - P_3$  (with the contents of the final register erased); the IGUS algorithmically generates a theory or 'schema' of the world based on the contents of the registers; and it algorithmically determines how to act based on the contents of the 'now' register  $P_0$  and the schema.

stores it in its registers; that's gathering. Then it has the capacity to operate algorithmically on the contents of its registers and its current inputs to determine a course of action implemented by its output devices (emit a beep or move away, for instance); that's utilizing. Hartle's (2008) IGUS is shown in figure 1, though few details are significant for most of what follows. The one simplifying assumption I will make is that the IGUS's registers are in some special 'empty' ready state before they are filled with gathered data; that assumption seems to be no restriction at all on the computational power of IGUSs, and anyway I believe that the points I wish to make could be made in a more complicated way without it. Clearly an IGUS can be realised as a digital computer, though the intent is that an IGUS has greater autonomy than the typical personal computer. A little more specifically, I want to consider an IGUS rather like us: especially responding to the same kinds of features of the world that we do, drawing inductive inferences from them<sup>7</sup>, and drawing on those inferences for action in the pursuit of specific goals. Below I shall say more about the operation of the IGUS, specifically its algorithm.

Insofar as the system has reliable procedures for forming veridical memories, insofar as it draws true inductive inferences, and insofar as its actions realise its goals on the basis of those inferences, I say that the IGUS models an important aspect of our knowledge. Computationally it is relevantly like us, and an IGUS could even be physically implemented like us. I trust it is clear enough and uncontroversial enough why I say that. For now I want to focus specifically on its memories. ('Memory' can be ambiguous between mnemonic knowledge and the registers of a Turing machine; I will always use it in the former sense.)

An IGUS 'remembers' by reading the records stored in its registers, so let's think about how that might go. One might initially wonder whether an IGUS could remember by literally reading records: explicitly deducing an intermediate event from knowledge of ready and later states of its registers. But it's easy to see that that would be a poor procedure to employ. Such an algorithm could form representations of the current state of a register (perhaps the register is its own representation in that regard), and *also* of the register's ready state, and then compute a representation of the cause of the difference. That third, derived, representation would then serve as a computational input for any decision making processes related to the memory (perhaps to report that the IGUS witnessed a particular event an hour ago). But what would the point of the computation be? A more efficient algorithm would simply assign to the contents of the register whatever computational role the derived representation plays in the algorithm discussed. See table 1 for a very simple example of how such a routine would work.

That is, the states of the registers themselves are perfectly capable of playing the computational role directly, without any interpretational process – the mnemonic knowledge is 'immediate'. But without some process such as I described, no *explicit* record reading of the kind considered in T&C has occurred, because no role has been played by knowing, or assuming the ready state. Since the algorithm employs no routine to deduce anything from representations of the present and ready states of its registers, the record reading is *implicit*.

Previously I emphasized that we shouldn't take T&C as giving a literal account of how we read records of the past. The observation just made shows how we (or at least the IGUS that models us) does it instead. In particular, the IGUS remembers without being subject to the record regress; there is no question of the IGUS having to compute the ready state of its registers by some reading some record of an earlier time, since the meaning of the registers' contents is implicit in its algorithm.<sup>8</sup> The IGUS shows that there is no paradox about memories, since it remembers without

<sup>7</sup>For recent work on algorithmic inductive reasoning see Waltz and Buchanan (2009).

<sup>8</sup>And if you think about it, for much the same reasons, even an explicitly record reading IGUS needn't explicitly



$P_0$	$P_1$	$C$
?	!	Yes
?		No

Table 1: A (partial) machine table for Hartle’s IGUS, specifying what is to be inscribed in  $C$  given the contents of  $P_0$  and  $P_1$ .  $P_1$  contains records of inputs one cycle ago; suppose that if the IGUS ‘saw’ an exclamation mark then  $P_1$  will now contain a ‘!’, and otherwise be left empty.  $P_0$  contains the current input; if the IGUS is asked whether it just witnessed an exclamation mark then suppose it will contains a ‘?’. The output depends on the contents of  $C$  (for instance, the IGUS will say ‘yes’ if  $C$  contains yes): this IGUS answers veridically. The point is that this IGUS ‘remembers’ whether or not it saw a ‘!’ without computation involving explicit representations of the  $P_1$ ’s ready and final states; the machine table depends directly on the present state.

knowing anything about the PH! Now, resolving the paradox doesn’t refute the position of T&C, which is that the PH plays a novel role, separate from the second law, in explaining the knowledge asymmetry – a conclusion argued for by a record regress argument. But now we’re absolutely clear that the regress does not apply literally to memories, we should ask seriously whether it applies at all. In the following sections I will use our IGUS to argue that it does not.

Before moving on, it’s probably worth justifying the claim that the IGUS ‘knows’ anything; in what sense can we call the contents of the registers ‘knowledge’? It will suffice here to point out that the IGUS can satisfy a wide range of behavioural criteria for knowing: actions depending on the memories, production of representations of the memories, communicating the memories to other IGUSs, for example. I do not mean to rely on a behaviouristic account of knowledge – these are just evidence, not constitutive.

## 2.2 An IGUS Remembers

Since our IGUS doesn’t explicitly use record reading, we should now ask whether it can have mnemonic knowledge that does not satisfy the necessary condition proposed in T&C – it doesn’t suffer from the record regress, which is the argument for the condition. To answer this question we have to say something more definite about the ‘presently surveyable condition’: what is entailed for its microstate by its remembering some event? Well, assuming that the IGUS works something like Hartle’s, depositing data representing particular inputs in identifiable registers, then a natural suggestion is that the memory of an event entails that the relevant registers be in the appropriate states. Let’s be much more generous than that: let’s include in the presently surveyable condition the states of all the registers (and its ‘schema’ if you like, though for brevity I won’t explicitly).<sup>9</sup> So to make sure we haven’t left anything out of their physical description, let the surveyable condition include the microstates of the registers.

Under these assumptions, however, the IGUS can have mnemonic knowledge that does not satisfy the necessary condition for knowledge of the record reading type – T&C’s condition is

---

read any records of times *before* the ready state obtained. The regress doesn’t apply to it either! But our IGUS is more natural, and since it doesn’t require any explicit record of the ready state, makes the point most straight-forwardly.

<sup>9</sup>If the IGUS is a universal Turing machine, then its algorithm will also be encoded in the registers. Include those registers too; just remember that mere marks on a tape (or whatever) aren’t an algorithm, *except in relation to some specific machine*, that acts on them in some specified ways. (I.e., one can’t infer a machine table from marks alone.) In other words, even in a UTM there is a distinction between the physical states of the registers and the algorithm.

*not* necessary. Just let the IGUS be the witness of the pachinko experiment discussed earlier, so that there are no macrotraces of the outcome, only the inscriptions in its registers. How might conditioning on these make the objects of the memories likely? Perhaps one can work backwards from the records to determine their causes. But that won't work, because any set of registers and any contents are compatible with many machine tables and initial states; hence the process by which the records were made cannot be determined. Nor does it help to also condition on the microstate; many different Turing machines could be realized in conditions with registers in identical microstates – the differences would lie in the physics of the rest of the device. In the present case then, any physical realization of any computational state, is equally compatible with the ice having fallen left or right (or any other possible event), so presumably conditioning on the state of the registers makes each outcome equally likely. And this even though the IGUS has a machine table along the lines of table 1, and the appropriate inscriptions, so that it does know in which beaker the ice fell. (And once again, the circumstances in which the only records of an event are in the brain are perfectly common: add to earlier examples of this kind of thing, remembering the result of a coin toss, or the exact spot the ball bounced, or who ate the last pretzel.)

But these considerations also show that our guess about what memories entail for the physical state of an IGUS is wrong. As we saw in the previous section, what register inscriptions mean, what they are memories of, depends on their computational role in an IGUS; for instance, that is what determines their effects on behavior. But of course the inscriptions in an IGUS's registers – or even their microstate – doesn't determine their computational role, again because any registers and contents are compatible with many Turing machines. So the surveyable condition under discussion is compatible with many computational roles, and hence with the register inscriptions being memories of many different things (and of nothing) – so we can't have cashed it out right, and we must try again.<sup>10</sup>

The next obvious thing to try is to include everything in the surveyable condition – the full microstate of the IGUS. A real, material IGUS is, computationally, nothing but a physical realization of a Turing machine, with no magical processing powers; hence its machine table should supervene on its physical state and the laws of physics. Hence we do not have the underdetermination problem raised above; the surveyable condition of an IGUS, in the new account, is not compatible with multiple algorithms. But is T&C's condition necessary with this new account? Are events remembered by the IGUS likely according to the Newtonian contraption, conditioning on the macrostate of the universe and the microstate of the IGUS (and laws and PH)? No.

Even though the machine table is now determined by the surveyable condition of the IGUS, working backwards to the causes of the inscriptions still won't work. Suppose we take the current microstate, turn the nomic handle in reverse, and try to evolve things backwards to earlier times, through changes to the registers, back to the states of the IGUS's input devices when they were triggered, to the stimuli they received, from which the causes of those stimuli – the recalled events – are (let's suppose) determinable.<sup>11</sup> The backwards determinism of the laws means that a closed system can be uniquely wound backwards from a given microstate (so there are no statistical assumptions here). But of course that won't help in this case because the IGUS is not a closed system: its state is the result of its initial state and a sequence of inputs (and outputs), which are

---

<sup>10</sup>I'd like to thank some very bright students at Oberlin College for pushing me on this point.

<sup>11</sup>If the IGUS works as we do to build up a coherent 'schema', then it will have routines to infer more from its inputs than is logically entailed by the raw physical stimuli by means of the laws. Various tricks and sleights of hand reveal the operation of such routines – which presumably are generally reliable under normal conditions. Let's shelve this complication.

nothing but interactions with the rest of the world (so the point holds even if the IGUS is otherwise perfectly screened off). Without those boundary conditions the laws do not determine the earlier states, but they are not fixed by the surveyable condition (nor could they be, without begging the question). Computationally the point is that, for many Turing machines, a given state of the registers can be compatible with many initial states and sequences of inputs.

(And note that there is no reason at all to think that conditioning on the PH will make some particular past history more likely: only if the PH alone entailed which of the possible stimuli were received would the problem now be soluble; but that would be tantamount to the PH alone entailing the causes of the stimuli, which is just the remembered past history. And no one thinks the PH does that!)

Given this new underdetermination, one might wonder how the IGUS manages to read the records in its registers, apparently beating physics! But there's no real puzzle: although the IGUS doesn't explicitly condition on the registers' ready state, it does do so implicitly. Consider, for instance, table 1: there we just supposed that the correct inscriptions would end up in  $P_1$ . But suppose, for instance, that the mechanism for inscribing the contents of  $P_1$  assumes that it is initially empty; if instead  $P_1$  starts with an inscribed '!' then the IGUS will end up saying 'yes' even if it did not see an exclamation mark. In other words, the IGUS's algorithm implicitly assumes an initial state for its registers for its proper functioning, and so is not merely conditioning on its current condition.

Now, there are Turing machines for which given initial and final states are compatible with many sequences of inputs, but a device with a machine table like that could not reliably recall the events that led to such underdetermined inscriptions. Suppose, for instance, that there are two possible inputs compatible with a given inscription and initial state. The computational role of the inscription could be that of a memory of one of the inputs, so that it functioned just like an IGUS memory of a specific event. But the underdetermination of the inscription's cause means that the device's algorithm cannot reliably determine which input produced it, so which input is apparently remembered is arbitrary. Under these conditions, even if the apparent memory of the device was, by chance, of the inscription's actual cause, we would not say that the device *knew* it, exactly because of arbitrary nature of the putative memory. In other words, an IGUS that models memory satisfactorily will not have a machine table of this sort; instead, it will have an algorithm such that the initial and current states of its registers are uniquely compatible with a series of inputs.

This new constraint on our IGUS's algorithm of course also constrains its physical constitution: it realize such a procedure. So, given this additional piece of information about the IGUS, will the current microstate in conjunction with backwards deterministic laws now entail the stimuli received by the receptors? (And hence, if everything works out, entail a high probability for their causes.) Of course not. There are algorithms for which the current *and* initial states determine the series of inputs, but for which the current state *alone* does not. Thus using the laws to time-reverse the evolution from the microstate of a physical realization of such an IGUS need not entail the prior inputs, let alone the stimuli that caused them. (In some particular realization it might, but the question is whether the necessary condition captures what is knowable in principle.) And this, even when the contents of the registers were properly formed and have the appropriate computational roles to entail that the IGUS knows that the events that caused them indeed occurred. So expanding the surveyable condition and constraining IGUS's machine tables is not enough to ensure that when an IGUS has a memory, evolving the state backwards will make the object of the memory likely. (And as before, the PH does not fix the problem.)

Some may be tempted to say that the 'memories' of an IGUS of this kind aren't really knowledge

either – after all, with respect to the different possible initial states, the inputs are ‘arbitrary’. But they are not arbitrary for the IGUS; after all, its algorithm relies for its proper functioning on the existence of some special initial condition, given which, we demand, the series of inputs *is* determined. So the underdetermination is crucially different from the previous one: if the initial and final states do not determine the series of inputs, then only chance can make apparent memories veridical; if the initial and final states do determine the inputs, then the memories will be reliable on the condition that the IGUS was indeed in the appropriate initial state. In the latter case, and only in that case, is it possible for the IGUS to work properly!<sup>12</sup> In my opinion the memories of such an IGUS constitute knowledge for it. To deny that, because it might have been in a different initial state than that it was (and so misread the registers) has the same logic as denying that I know anything from experience, even in the case that the experiences *are* properly produced by their objects, just because identical experiences *could* also have been caused by an evil demon. I know that some do adopt such positions, but it would be at least unfortunate if one were forced to in order to defend the view of T&C. I would especially note that one would reasonably expect there to be considerable computational efficiencies to be had by our kind of IGUS over one whose current state alone determined the sequence of inputs. Hence it is plausible that evolution made us along the lines of our kind of IGUS, and hence any argument that it doesn’t have knowledge of the past plausibly applies to us as well – some might accept that conclusion, but it seems a high price to pay.

Let’s put such skepticism aside then, and take it that an IGUS can have knowledge of the past even though the conjunction of its algorithm and current computational state doesn’t determine its sequence of inputs (though the additional conjunction with the initial computational state does). Then there is nothing in principle preventing the construction of such a device for which the same is true of the laws governing it and its microstate: they don’t determine the sequence of inputs (although in conjunction with the initial microstate they do) – even though the events that caused those inputs are known to the IGUS. And when those events are like the outcome of the ice pachinko experiments, so that conditioning on the macrostate of the universe doesn’t make them probable either, then the full Newtonian contraption fails to make them probable, even though they are perfectly well known to the IGUS – showing that *we can know more of the past than the contraption allows*.

Note that conditioning on the PH makes it unlikely that the IGUS and its current state is the result of a fluctuation from equilibrium, but the PH does not make it probable that the IGUS’s memory is veridical, because that further requires that the IGUS started in the correct initial computational state, out of all the many possibilities. The kind of macroscopic information that the PH provides about the very distant past cannot – as far as I can see – make that likely.<sup>13</sup>

---

<sup>12</sup>Of course I’m making the usual background assumption that the IGUS is constructed so the computational roles of the inscriptions are commensurate with the inputs they record; the point is just that such a thing is possible because, given the appropriate initial state, the inscription is unequivocal about its cause.

<sup>13</sup>David has put it to me that the contraption will assign a high probability to any IGUS one finds being the product (direct or indirect) of natural selection. He further argues, I take it, that selection produces fit systems, so it is thus also likely that any IGUS has veridical memories; hence the contraption must, after all, assign a high probability to the correct IGUS initial state and the objects of its memories. I have two main responses. First, not all IGUSes are reliable (indeed, some IGUSes have veridical memories of the unreliability of others, and one can envision situations in which only unreliable IGUSes survive some cataclysm): in short, a blanket appeal to natural selection threatens to include too many mnemonic systems as reliable. Second, I think David has the view that if the contraption doesn’t make *our* memories likely veridical, then we should doubt them, and fall into skepticism. But that worry is misplaced if our memories do exceed the reach of the contraption, as I have argued. For in that case the contraption simply fails to condition on all the relevant information: on the PH and the presently surveyable

Nor can there be some way ‘better’ way to calculate the probabilities, in which they turn out to be different, since the approach here already conditions on *everything* available to the contraption: it doesn’t condition on anything but the PH, the surveyable condition and the laws, and all of those were considered. Since probabilities are unequivocal, any calculation that conditions on the same things must produce the same result. For instance, one might have entertained the following idea: the IGUS’s microstate determines its machine table and the states of its registers, so it determines their computational roles, and so, we have supposed, their intensional contents – what they are memories of. But shouldn’t the IGUS remembering X imply that X is likely (given the PH); and since the IGUS is physical, the probabilities must be physical, so surely those of the Newtonian contraption? But that, we have just seen, is wrong – recalled events need not have high conditional probabilities. (Of course one could modify the PH to include more information about the initial microstate; since the underlying laws are deterministic, conditioning on enough initial information will entail anything you like about later states. But the claim that the contraption *with some hypothesis or other* about the past assigns correct probabilities is, given determinism, empirically irrefutable: it is considerably weaker, and less interesting than the claim of T&C.<sup>14</sup>)

The issue we are touching on here is how to understand the physical basis of intensionality in a way that connects beliefs up to their physical causes. It isn’t one in which I have any confidence in my intuitions, so I don’t really understand how such a calculation would actually proceed – how to present it other than in the sort of hand-waving way I just did. But I will make the following observations: as we have now seen at length, in the IGUS model, its current microstate determines the content of its memories, but their objects need not have high probabilities conditioned on the microstate. So if we think that the object of a memory ought to have a high physical probability, then that must require conditioning on something more, some physical background assumption. I suppose that one could further condition on the IGUS’s memories being likely true (our recent discussion of course shows that doesn’t necessarily follow from the surveyable state), but that doesn’t seem very illuminating. Another alternative is to also condition on the initial microstate of the IGUS; that would at least determine the sequence of inputs, so perhaps could determine something about the stimuli that produced them, and so on.

But the bottom line is that an IGUS’s knowledge of the past is not constrained by the necessary condition, and so its subjective probabilities need not agree with the probabilities of the Newtonian contraption.

### 2.3 Final Thoughts

You may have noticed that this paper has not addressed the knowledge asymmetry directly; instead the focus has been on how we read records. Of course it’s a short step to the asymmetry, for it apparently lies in the existence of readable records of the past but not of the future. None of my remarks are intended to express skepticism that the PH is relevant to the explanation of the knowledge asymmetry, perhaps even in ways separate from its role in the second law. For instance, brains and computers we know how to build are highly time directed systems: synapses aren’t time-reversible, and neither are logic gates. The processes in these systems are governed by cause and effect, so perhaps the arrow of causation lies at the heart of the knowledge asymmetry – and

---

state, but not on our memories. These points are not knock-down arguments against the natural selection argument, but show that it requires considerable articulation and independent argument. (This conclusion is weaker than that in main text because it was added at the editorial stage, after David had written his response to the stronger statements.)

<sup>14</sup>This comment was added after David had the chance to respond.

the PH at the heart of the causal asymmetry? (This problem is related to the difficulty of building an IGUS with memories of the future – if future events can't influence it now, it seems the maker just has to fill the registers himself!)

Hartle suggests that his IGUS (figure 1) is time asymmetric because its operation involves erasing data from the final register, and because the permanent loss of such information is entropy increasing; thus the system can only work in one temporal direction in virtue of the second law. I'd like to point out that this proposal can't be right as stated: for instance, in his IGUS the final register remains empty until the fourth cycle, so it isn't erased for that initial period. So he hasn't explained the knowledge asymmetry until that time; by his lights, nothing rules out a suitable IGUS from remembering the future *until* all its last register was inscribed, and erasing had to commence! To apply Hartle's idea, some more general argument about the necessity of erasing would be required. For instance, doesn't shuffling the records down the registers involve erasing? But that's not an essential feature. Suppose, for example, that the registers are buckets and data are stored by placing given numbers of balls in them: shuffling could be carried out by tipping the balls from one bucket to the next. The sequence of records inscribed by the IGUS's detectors can easily be inferred from the collection of registers (by subtraction, of course). The problem is that this or that part of the IGUS can be constructed without erasing, so something quite general about the physical implementation of Turing machines would be needed to pursue Hartle's strategy.

One final observation. Consider knowledge of the past that we draw from records other than memories. Clearly we do read records in the way described in T&C all the time: the footprints in the snow, the dent in the bumper, the puddle in the street. We know what they signify because we infer what caused the changes from a 'ready state', even if we don't remember the causes – and even when we have no memory of the ready states (perhaps I've never been down this street before). Does a record regress still threaten here? No.

We can easily imagine an IGUS that can read the records because it possesses *generalizations* about the ready states of systems, generalizations that it obtains by induction on the contents of its memories. (It's for this reason that I gave it inductive powers when I introduced it.) Significantly, true generalizations, if informative at all, are often more informative about origins than fates: all acorns start much the same, as do all sandy beaches as the receding tide exposes them, and all rocks – at least compared to the countless ways in which acorns can develop or not into oaks, the different kinds of impressions that can be made on beaches, and the different kinds of marks, including those made by humans, that might be found on a rock. And similarly for freshly fallen snow, bumpers and streets. (This point is emphasized by Lockwood, 2005, Chapter 11.) That is to say there are such records of the past because one can infer an earlier ready state simply from the kind of system involved; but that there are no corresponding records of the future, because there are no generic later states for a given kind of system. Thus, we would have an explanation of the knowledge asymmetry for the inferred (rather than remembered past), if we could only characterize and explain this initial-final state asymmetry!

I won't address that here, but note that we finally have a model of our knowledge of other times. There are memories, modeled by the IGUS, in which the meaning of records is determined by their computational role. Using these, and a capacity for induction, we can learn to read other records for what they tell us of the past. This model allows a response to T&C's objection (119) that the brain's ready state can't be the 'mother' of all ready conditions: that memories are defeasible by external records. There's nothing odd about that in the present model, we simply have a set of conflicting beliefs: some memories and some concerning current experiences, and some inductive beliefs, which generalize other memories. Something has to give, and whether it is a specific memory

or a generalization is going to depend on our determination to hold onto the different beliefs.

### References

Albert, D. Z (2000) *Time and Chance*: Harvard University Press, Cambridge, MA.

Gell-Mann, M. (1994) *The Quark and the Jaguar*: W. Freeman, San Francisco.

Hartle, J. (2008), "The Physics of Now", *Am. J. Phys.*, 73, 101-109 (2005), gr-qc/0403001.

Lockwood, M. (2005), *The Labyrinth of Time*: OUP, Oxford.

Waltz and Buchanan (2009), *Science*, 324 no. 5923, pp. 43 - 44.