Oaksford, Michael and Chater, N. (2014) Probabilistic single function dual process theory and logic programming as approaches to non-monotonicity in human vs. artificial reasoning. *Thinking & Reasoning* 20 (2), pp. 269-295. ISSN 1354-6783.

Probabilistic Single Function Dual Process Theory and Logic Programming as Approaches to Non-monotonicity in Human vs. Artificial Reasoning

Mike Oaksford*

Birkbeck College, University of London

Nick Chater

Warwick Business School, University of Warwick

Short Title: Dual processes in non-monotonic reasoning

*Corresponding author:

Mike Oaksford

Department of Psychological Sciences

Birkbeck College

University of London

London, WC1E 7HX, UK

e-mail: Mike.Oaksford@bbk.ac.uk

telephone: +44 (0) 20 7079 0879

Acknowledgement: We thank Jonathan Evans, Ruth Byrne and an anonymous reviewer for their highly constructive criticisms of the first version of this paper. Their diligent reading motivated many important revisions.

Abstract

In this paper, it is argued that single function dual process theory is a more credible psychological account of non-monotonicity in human conditional reasoning than recent attempts to apply logic programming (LP) approaches in artificial intelligence to these data. LP is introduced and among other critiques, it is argued that it is psychologically unrealistic in a similar way to hash coding in the classicism vs connectionism debate. Second, it is argued that causal Bayes nets provide a framework for modelling probabilistic conditional inference in System 2 that can deal with patterns of inference LP cannot. Third, we offer some speculations on how the cognitive system may avoid problems for System 1 identified by Fodor in 1983. We conclude that while many problems remain, the probabilistic single function dual processing theory is to be preferred over LP as an account of the non-monotonicity of human reasoning.

Dual process theories (Evans 2003, 2007; Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996; Stanovich, 2011; Stanovich & West, 2000; Wason & Evans, 1975) invoke two separate cognitive systems to explain performance on a variety of cognitive tasks. These are labelled System 1 and System 2. System 1 is rapid, parallel, automatic, do not require the resources of working memory (WM), and only their final product is posted in consciousness. In contrast, System 2 is slow sequential, and analytic and makes use of the central working memory system. In particular, System 2 ‘‘permits abstract hypothetical thinking that cannot be achieved by System 1’’ (Evans 2003, p. 454). Recently, Oaksford and Chater (2012) argued that accounting for non-monotonic or defeasible reasoning in dual process theory required that both System 2 WM representations and System 1 long term memory (LTM) representations need to be interpreted probabilistically. This position is consistent with Evans and Over (2004; see also Over, Evans, & Elqayam, 2010) adoption of probability logic (Adams, 1998) as underpinning analytic processes in System 2. But it contrasts with accounts which treat analytic processes in System 2 as underpinned by standard binary truth functional logic (Heit & Rotello, 2010; Klauer et al., 2010; Rips, 2001, 2002; Stanovich & West, 2000; Stanovich, 2011). Oaksford and Chater (2012) labelled the former approach, the *single function dual process* (SFDP) approach and the latter the *dual function dual process* (DFDP) approach. Our goal in this paper is to confront some further problems and challenges for the probabilistic SFDP approach but first we rehearse Oaksford and Chater’s (2012) argument in detail.

Probabilistic Single Function Dual Process Theory

Both the dual process theory (Evans 2002) and the probabilistic approach (Oaksford and Chater 1991, 1998, 2001, 2007) developed out of a critique of the classical logicist approach to cognitive architecture (Fodor 1975; Pylyshyn 1984), which is a *logical* single function dual process theory. The store of world knowledge in LTM consists of a consistent set of logical formulae in the language of thought that can be combined with given information in WM using logical inference rules to yield new information in a proof theoretic derivation. Evans (2002) and Oaksford and Chater (1991, 2007) argued that the defeasiblity of human reasoning argued strongly against this logicist single function view. Defeasible reasoning creates two problems for such systems. In standard logic, defeasible reasoning leads to contradictions. Suppose that *if x is a bird then x flies* is part of your world knowledge in LTM, then when someone asserts that *Tweety is a bird*, you may validly infer that *Tweety can fly* and so add this to your world knowledge in LTM. But if you are then told that *Tweety is an Ostrich* your belief that *ostriches can't fly* will lead you to add *Tweety cannot fly* to your world knowledge in LTM resulting in a contradiction, i.e., *Tweety can fly* ∧ *Tweety cannot fly* ("∧" = and). One attempt to avoid this unacceptable conclusion is to propose a non-monotonic logic (Reiter 1985). However, as Oaksford and Chater (1991) argued based on critiques in artificial intelligence (McDermott 1987), that this leads to triviality—all that can be concluded is that *Tweety can fly* ∨ *Tweety cannot fly* (a tautology and something you knew before drawing any inferences, "∨" = or)—and to computational intractability, i.e., the Frame Problem (see, Oaksford and Chater 1991, 2007). Can a dual function view address these problems? Oaksford and Chater (2009, 2011) argued that it may not because the two systems must interact. But if the systems obey fundamentally different principles, it is not clear how this is possible.

Consider again the familiar example of inferring that *Tweety flies* from the general claim that *birds fly* and the fact that *Tweety is a bird*. On the DFDP view, this inference could be drawn logically in System 2 from the premises, on the assumption that birds fly is a true universal generalization; System 1, by contrast, might tentatively draw this conclusion by defeasible, associative processes, drawing on general knowledge. But a lack of synchrony between the two systems, presumed to operate by different rational standards, threatens to cause inferential chaos. Consider, for example, what happens if we consider the possibility that *Tweety is an ostrich*. If System 2 works according to logical principles, the clash of two rules threatens contradiction: we know that *birds fly*, but that *ostriches do not*. To escape contradiction, one of the premises must be rejected: most naturally, *birds fly* will be rejected as false. But we now have two unpalatable possibilities. On the one hand, suppose that this retraction is not transferred to general knowledge and hence is not assimilated by System 1. Then the two systems will have contradictory beliefs. Moreover, if System 2 reasoning cannot modify general knowledge in System 1, its purpose seems unclear. On the other hand, if *birds fly* is retracted from world knowledge, along with other defeasible generalizations, then almost all of general knowledge will be stripped away—as generalizations outside mathematics are typically defeasible (Oaksford & Chater, 2007, 2009)—leading System 1 into inferential paralysis.

Oaksford and Chater (2012) argued that the best way to avoid these unpalatable conclusions and account for the defeasibility of human reasoning is to adopt the SFDP view in which representing *birds fly* in WM amounts to the assumption that the probability that something flies given it is a bird is very close to 1 ($\Pr(\mathit{flies}(x)|\mathit{bird}(x)) \approx 1$). Consequently, rather than having to reject *birds fly* as false in System 2, the observation that *Tweety is an Ostrich* simply provides a negative instance that leads to a reduction of $\Pr(\mathit{flies}(x)|\mathit{bird}(x))$ in System 1 (or the inclusion of a defeater, see below). That is, the two systems can properly communicate. Oaksford and Chater (2012) argued that this position successfully accounted for a range of findings that had motivated dual process theories. For example, people do make non-modal responses apparently not explained by probability theory and these non-modal responses correlate with IQ (Stanovich & West, 2000; Stanovich, 2011). However, recently logic programming approaches to non-monotonic reasoning in Artificial Intelligence have been applied to human reasoning (Stenning and van Lambalgen, 2005). This application questions part of the motivation for the probabilistic approach and in this paper we address whether the logic programming approach provides an adequate account of actual human reasoning.

Logic Programming and the Probabilistic SFDP Approach

The arguments for a probabilistic SFDP approach were formulated with Reiter's (1985) default logic in mind, where rule application involves intractable consistency checking between the conclusion of an inference in System 2 and world knowledge in System 1. However, recently Stenning and van Lambalgen (2005) have proposed that a different *Logic Programming* (LP) approach to non-monotonic reasoning which can address the problems of maintaining consistency within and between Systems 1 and 2. In this paper, we argue that single function dual process theory is a more credible psychological account of non-monotonicity in human reasoning than this attempt to apply logic programming approaches in artificial reasoning to the human data. In addressing some of the psychological evidence, LP appears to hold out the promise of a *local* computational theory of non-monotonic reasoning, i.e., one that appeals only to the premises in System 2, and which does not engage world knowledge in System 1. However, we argue that this is illusory and that to generalise beyond the single experimental paradigm to which LP has been applied will involve more *global* processes involving System 1. Moreover, we argue that the mechanism by which LP renders such global processing computationally tractable in System 1 is unlikely to be psychologically real.

An important element of the argument against LP is that there are various inferences that are naturally accounted for in the probabilistic SFDP approach that cannot be explained in the LP approach. In establishing this point we will present some arguments that causal

Bayes nets (CBNs) can provide a good framework within which to develop a theory of conditional reasoning and we show that CBNs can naturally account for these patterns of inference (Ali, Chater, & Oaksford, 2011; Fernbach & Erb, 2013; Oaksford & Chater, 2007, 2013).

The argument for a probabilistic SFDP theory relied on the need to maintain consistency between System 1 and System 2, i.e., between the representations being manipulated in WM and the relevant representations in LTM. As we just observed, LP seems to provide a way of "maintaining consistency *within* and between Systems 1 and 2." That is, it goes beyond our arguments for probabilistic SFDP. While we believe the cognitive system capable of maintaining local consistency in System 2 and between System 2 and *relevant* parts of System1, we doubt that the cognitive system can maintain the global consistency of System 1. LP uses a form of indexing to ensure the consistency of System 1 which also provides for tractable search over world knowledge for possible exceptions, e.g., like Tweety is an Ostrich. In criticising LP, we question the psychologically reality of this indexing scheme. While lacking psychological reality, this scheme nonetheless does help solve the technical *frame problem* for classical Artificial Intelligence (Shanahan, 1997). However, there remains the epistemological frame problem that one cannot circumscribe the information in System 1 that is *relevant* to any particular inferential goal being pursued by System 2 (Fodor, 1983, 2001). Any information is potentially relevant, a point that Fodor (1983) labelled "isotropy." Computing relevance, which is global, open-ended and context sensitive, again seems intractable. In the final section, we offer some, highly preliminary speculations on how in everyday situations people may avoid these problems in part by appeal to an alternative philosophy of science to that implicit in Fodor's arguments.

In summary, in this paper our goal is to argue that single function dual process theory is a more credible psychological account of non-monotonicity in human reasoning than attempts to apply LP approaches in artificial intelligence to these data. To achieve this goal, we first provide a substantive critique of LP, which might otherwise be thought to provide an alternative to a probabilistic SFDP theory. Second, we show that CBNs provide a good framework for modelling conditional inference in System 2 because they can deal with patterns of inference LP cannot. Third, we offer some speculations on how the cognitive system may avoid problems for System 1 created by the putative need to compute properties like relevance and plausibility. We first outline the LP approach.

Logic Programming

Recently the negative implications of defeasible reasoning have been questioned (Stenning & van Lambalgen, 2005; Kowalski, 2010). In Stenning and van Lambalgen's (2005) approach, rather than employing the M-operator, they propose a different account based on logic programming which seems to hold out the promise of avoiding intractable computations in System 1.

Stenning and van Lambalgen (2005)

There are four important features of Stenning and van Lambalgen's (2005) theory. First, it draws a distinction between *credulous* interpretative reasoning and *sceptical* critical reasoning. The former uses a weak logic in order to infer an interpretation of premises in which they are true. This is a computational embodiment of Davidson's (1974) *principle of charity* in language interpretation. *Sceptical* inference involves critically examining the truth or falsity of the premises/utterances employing standard binary truth functional logic. Second, in the propositional case, conditionals are interpreted as always having a complex conjunctive ($\wedge$) antecedent with the explicitly stated condition as one conjunct and the negation of an *abnormality* proposition (*ab*) as an implicit conjunct, e.g., where p = *x is a* *bird* and $q = x$ *flies,* "birds fly" would represented as, *if* $p \wedge \neg ab_1$ *then* q.[1] Third, *closed world reasoning* then serves to head off intractable global computations involving the whole of the contents of long term memory. In closed world reasoning, unless there is some statement that implies ab_1 in the data base ($\vartheta \rightarrow ab_1$) it is treated as false ($\perp \rightarrow ab_1$). This approach was first proposed in McCarthy's (e.g., 1986) account of *circumscription*. In logic programming, the analogous process is called *completion* or *minimization*, in which the closed world assumption removes negated items and the resulting *minimal model* only represents what is true. Fourth, the final wrinkle is that the connectives are interpreted in Kleene's strong three valued logic (Haack, 1974), in which a third truth "value," u, stands for undecided. In this system, it seems as if rule application does not involve global, intractable computations but merely a local presumption of normality. That is, until you learn more about Tweety it is safe to assume she is not abnormal.[2]

Using rules formulated in this way permits the logic to deal with the standard case as well as when a defeater is available using only local computation. We illustrate this using the suppression effect (Byrne, 1989). Take the following conditionals and categorical premises used in Byrne (1989) but labelled similarly to Wernhard (2011, p. 13):

C_p: *If she has an essay to write (p) she will study late in the library (q).*
C_s: *If she has a textbook to read (s) she will study late in the library (q).*
C_r: *If the library stays open (r) she will study late in the library (q).*
p: *She has an essay to write.*
q: *She will study late in the library.*
r: *The library stays open.*
s: *She has textbooks to read.*

[1] The reason for the subscript is that abnormality propositions must be indexed to particular conditionals, which define in what respect a proposition or object is abnormal. This means that there are a great number of distinct abnormality propositions/predicates that potentially need to be stored and possibility accessed. Although including these explicit labels does produce search results within computationally tractable bounds, empirically not just theoretically (for some empirical tractability results see, e.g., Grégoire, Mazure, & Saïs, 1998).
[2] This approach has been illustrated here and exampled with respect to empirical data (Stenning & van Lambalgen, 2005) only using propositional LP. However, to deal with the Tweety case, where the conditional is a generalisation, abnormality *predicates* would need to be used as in McCarthy (1986).

We first look at the straightforward modus ponens inference, involving just the premises C_p and p. They are interpreted in terms of the logic program in A.

A $\{p; p \wedge \neg ab \ \rightarrow q; \bot \rightarrow ab\}$ (logic program)

$\{p; p \wedge \neg ab \ \leftrightarrow q; \bot \leftrightarrow ab\}$ (completion)

$\{p; p \leftrightarrow q\}$ $\therefore q$ (minimal model)

The "$\bot \rightarrow ab$" clause embodies the closed world assumption that the abnormality propositional is assumed to be false ($\bot$). If a further conditional premise, C_r, is added, then the situation becomes as in B, where the negation of the antecedent of C_r, the library does not stay open ($\neg r$), now functions as grounds to infer that *ab* is true ($\neg r \rightarrow ab$).

B $\{p; p \wedge \neg ab \ \rightarrow q; r \wedge \neg ab' \rightarrow q; \bot \ \rightarrow ab; \bot \ \rightarrow ab'; \neg r \rightarrow ab; b'; \neg p \rightarrow ab'\}$

$\{p; (p \wedge \neg ab) \wedge (r \wedge \neg ab') \leftrightarrow q; (\bot \ \vee \neg r) \leftrightarrow ab; (\bot \ \vee \neg p) \leftrightarrow ab'\}$

$\{p; (p \wedge r) \leftrightarrow q\}$ q? (need info about r)

So in A, the interpretation arrived at allows the inference to *q*, whereas in B it does not and nothing can be inferred without further information about the status of *r*. For MP, this behaviour is close to that observed by Byrne (1989), although Stenning and van Lambalgen (2005) did not attempt to fit this logical model to Byrne's data.

Problems for LP

We now present some potential problems for the LP approach. We argue that it requires a dual process approach, i.e., some System 1 global computation is required (*Dual processes*), it is implausible as a psychological theory (*Psychological reality*), and it cannot capture certain patterns of inference that the probabilistic approach can handle (*The probabilistic approach*).

Dual processes

In this section, we consider four problems which seem to require that LP says something more about System 1 processing. First, we consider how different reasoning systems handle inconsistency. Second, we suggest the LP, like other theories, needs to address the computation of relevance in System 1. Third, conditions like *r* and *s* differ in important respects. i.e., one is an enabler the other an alternative cause. Determining this distinction needs to invoke general knowledge in System 1. Finally, we observe that the empirical fact that similar effects are observed in the implicit suppression paradigm (e.g., Cummins, 1995) requires System 1 to be invoked.

Inconsistency. Consider the standard single rule case and what happens under inconsistency, e.g., you believe that birds fly, that Tweety is a bird, but that Tweety cannot fly. Standard logical approaches, like mental logic or mental models, address the resulting inconsistency by claiming that the conclusion or one the premises are false. The LP approach provides *a second way* that does not involve revising our beliefs like this, i.e., these beliefs remain consistent as long as we now believe that the implicit conjunct in the antecedent, $\neg ab_1$, is false and so Tweety is abnormal with respect to flying. This is consistent with some major examples in the philosophy of science, our prototypical rational activity. So for example, when in 1781 Herschel observed pertubations in the orbit of Uranus, not predicted by Newtonian celestial mechanics, astronomers did not reject the theory. Rather they rejected the normality clause (what Putnam [1974] called *auxiliary assumptions*) that there were not more than seven planets. Couch Adams and Le Verrier, inferred there was an unknown eighth planet exerting a gravitational force on Uranus that could explain the pertubations and Neptune was finally observed 65 years later by Galle in 1846. The perihelion of Mercury, of course, proved less amenable to being *explained away* as abnormal in some respect. This anomaly for Newton's celestial mechanics eventually required its rejection using sceptical reasoning. It is worth noting, however, that this conclusion did not gain acceptance until an alternative theory, relativity theory, became available.

However, in LP just concluding that ab_1 is true, i.e., Tweety is abnormal with respect to flying, is a rather unsatisfying conclusion. We would want to know whether we had any good reason to believe this and this will involve searching LTM, i.e., System 1, to find some proposition ϑ such that $\vartheta \rightarrow ab_1$ and then checking the world to see whether ϑ is true of Tweety. This is the inferential process underlying explaining away the orbit of Uranus counterexample we just discussed. This need to search LTM of course implies LP is a dual process theory invoking both local System 2 and global System 1 computations. This is not in itself problematic as LP was explicitly designed to improve the tractability of such searches over LTM for world knowledge, i.e., System 1. By providing explicit indices for abnormality conditions, rather than inferring they do not exist from prior knowledge, such searches can be kept within tractable bounds over reasonably large data bases (see, footnote 1).

Relevance. However, it seems implausible to assume that every r_i such that $r_i \rightarrow ab_1$ is accessed in searching for good reasons to infer that Tweety is abnormal with respect to flying. Or indeed that just the first r_i accessed is brought to mind. Rather our strong intuition is that the most *plausible* or *relevant* r_i is what will come to mind. Of course, this is to appeal to global properties of the belief system that it is not clear that indexing defaults resolves. Why in the system is one abnormality condition r_i more plausible than another? Context could clearly disambiguate this inference. For example, if you are at a zoo, then $ostrich \rightarrow ab_1$ would seem more plausible than $broken\ wing \rightarrow ab_1$ but in the vets surgery (in the UK at least), these plausibility judgements would reverse. Appealing to context is just to label a major lacuna in theories of human reasoning and language interpretation (Miller, 1996) rather than offering a solution (although we will suggest in the final section that the deictic context may provide cues that may help avoid complex computations). Consequently, the LP approach does not vitiate the need for more global computations or offer a straightforward solution.

Disabler or alternative cause? In modelling the suppression effect (Byrne, 1989), LP also seems to require general knowledge to be accessed to determine the status of the antecedents of a conditional. Looking again at the conditionals used in Byrne (1989) in the section *Stenning and van Lambalgen (2005)*, Wernhard (2011, p. 15) states one important step in applying LP as follows:

> "If there are two conditionals with the same conclusion, determine whether the premise of the second conditional is an alternative to the first one, like s in C_s which is an alternative to p in C_p for concluding q, or is additional to the first one, like r in C_r. *This step requires to take* [sic] *contextual information and background knowledge into account.*"

So simply disambiguating the status of the second conditional premise, C_r or C_s, with respect to C_p will require further access to "contextual information and background knowledge," i.e., seemingly to non-local System 1 processes.

The implicit suppression paradigm. Cummins' *implicit* suppression paradigm also raises an empirical problem for LP's ability to maintain relatively straightforward local System 2 computations in modelling the suppression effect, (Cummins, 1995; Cummins, Lubarts, Alksnis, & Rist, 1991; Sellen, Oaksford, & Gray, 2005). In these experiments, participants only ever see a single conditional premise, i.e., C_p, no explicit information is given regarding C_r or C_s. However, each of the C_p used were pretested for the number of alternative or additional antecedents they allowed. In the inference task, with different participants, almost identical effects were observed as in the Byrne explicit suppression paradigm, indicating that information similar to that in C_r or C_s was being accessed from

LTM for world knowledge, i.e., from System 1. That is, people seem to be spontaneously accessing *plausible* abnormality conditions which affect their inferences in just the same way as if they were explicitly present. This is the paradigm that has come to dominate research on conditional inference over the last 20 odd years and it does not seem amenable to a straightforward local System 2 approach.

In summary, it would appear that the LP approach is only able to maintain a local approach to explicit suppression tasks. When we move to belief revision in the face of apparent inconsistency and to the implicit suppression task, there seems to be a need for more global System 1 processes. Some of these problems are shared with other approaches like Causal Bayes Nets. Consequently, it is important not leave this section with the impression that other approaches can fully resolve all of these problems. However, the next set of problems, are more discriminatory between the probabilistic SFDP and LP approaches.

Psychological Reality.

When taking up ideas developed in artificial intelligence (AI) for use as psychological theories it is a good idea to have a reality check on whether they are plausible, that is, to pose the question, are they likely to be psychologically real? Of course, in AI it is perfectly acceptable to come up with neat ways of making a process like non-monotonic reasoning computationally tractable. And here using abnormality propositions as in LP, has proved to be a very valuable tool and much better than a kluge, i.e., a cheap, non-generalizable fix. However, how plausible is it to propose that people are explicitly indexing exceptions as abnormal? In this section, we raise three problems for this approach. First, indexing defaults is similar to hash coding as a way of implementing content addressable memory and seems equally implausible as a psychological theory. Second, the approach to learning and inference implicit in the LP approach seems to be inconsistent with the large body of evidence for the current division between System 1 and System 2 processes. Finally, the focus on abnormality as a way of encoding disablers suggests that natural language should be replete with imprecise expressions capturing these conditions but this does not seem to be the case.

Learning, abnormality and hash coding. In a practical data base, every conditional statement that is defeasible would require its own abnormality proposition/predicate, ab_i. Furthermore, further clauses would need to added showing which actual defeaters lead to abnormality, $r_j \rightarrow ab_i$. So people have to learn what is normal and what is abnormal and explicitly index the abnormal cases. This proposal is redolent of the old debate in connectionism (Chater & Oaksford, 1990; Rumelhart & McClelland, 1986) about hash coding as an approach to content addressable memory. Hash codes assigned a unique index to each memory location for the descriptors of an item. So if one descriptor is presented, others with the same hash code can be rapidly accessed. But to do this requires all the combinations of descriptors of an item to be known in advance. In the connectionism debate, the question posed was, well yes, the cognitive system could do it this way but is it likely? The answer was negative because the cognitive system has to *learn* what goes with what in setting up a content addressable memory and this requires learning the co-occurrence statistics of descriptors from our interactions with the world.

Similarly, in LP what is normal has to be learnt and again it has to be learnt from the statistical structure of the world. We only come to know that birds normally fly because most of the birds we have observed can fly. The existence of exceptions ensures that this is a statistical norm, i.e., the probability of x flying (q) given x is a bird (p) is high. Those exemplars that provide the proportion in the p, $\neg q$ cell are the exceptions. In J. L. Austin's (1960) terms, the concept of normality is the "trouser concept" and it is a statistical one that we must learn from the world. We are not handed the normal and abnormal cases and just have make sure that they are properly labelled.

System 1 vs system 2. So the LP approach trades the complexity of learning, i.e., filling the data base in System 1 with knowledge including knowledge of abnormality conditions, against the complexity of inference in System 2. Learning is a lengthy, complex, and slow process, whereas inference is perhaps fast and local.

There are two problems. First, no account is provided of how labelling conditions as abnormal is incorporated in to the learning process. Second, such an account may not locate the complexity profile in the right place to match up with dual process theory in which rule application and inference is slow and effortful albeit also local (Evans, 2007; Stanovich, 2011). This is because System 2 analytic processes involve effortful reflective thought, considering alternative possibilities and perhaps adopting a sceptical approach (Evans, 2007; Stanovich, 2011). However, Stenning and van Lambalgen (2005) argue for an efficient implementation of minimal models in neural networks over which inference is as rapid as propagating activation from one level of the network to another. This formulation implies that inference over the model in working memory, i.e., System 2 processing, is rapid and non-effortful which is not consistent with current dual process theories. In addressing this mismatch, all Stenning and van Lambalgen (2005, p. 954) say is that they draw the boundary between System 1 and 2 in a different place without addressing the large body of evidence cited by Evans (2007) and Stanovich (2011) for its current location.

A possible resolution is to regard LP as about interpretative processes, i.e., *reasoning to an interpretation*. These processes are likely to be a quite rapid System 1 processes, like language comprehension and inference over neural networks generally. But even then the identification of System 2 with a minimal model represented as a neural network is incompatible with current dual process theory.

Abnormality and natural language. If we always had to index conditional knowledge with abnormality propositions or predicates one might expect natural language to be replete with imprecise expressions capturing these conditions. However, other than "broken" and "not working", English does not seem to contain a large body of terms to describe abnormality or a perhaps an affix signifying abnormality. For example, if you believe that *if you turn the key, the car starts* and you therefore *turn the key* but *the car does not start* you would most likely describe this situation as *the car did not start*. It seems unlikely, that one would articulate the denial of the condition that makes the situation non-contradictory given the default rule interpretation of this conditional. Of course, this may happen and people who know very little about cars may say *the car did not start*, *it must be broken*. But, we suspect they are more likely to say, *the car did not start*, *did you refuel last night* or *did you leave the lights on…etc.* where failures of these conditions ($\neg r$) are defeaters. Although it is possible that getting to these conclusions are mediated by "silent" abnormality propositions, i.e., $(p \wedge \neg ab_1) \rightarrow q, p, \neg q \therefore\ ab_1, \neg r\ \rightarrow\ ab_1, \therefore \neg r$ (the latter inference is a data base query, ab_1?, i.e., given ab_1 is true what else needs to be true). Abnormality propositions/predicates while being a useful way of providing more tractable non-monotonic data-bases in AI, do not seem to leave much of a trace in our everyday way of speaking about these situations.[3]

In summary, LP and abnormality propositions/predicates have proved to be a very important tool in addressing non-monotonic inference and the technical frame problem in logic programming in AI. However, like hash coding and content addressable memory, we think it unlikely that this is how the human mind solves this problem.

The Probabilistic Approach

In this section, we argue that there are important inferences that can be captured by the *probabilistic* SFDP approach but not by LP. The new paradigm in reasoning (Manktelow,

[3] Although one could argue that the prefix "ver" in German may fulfil something like this role (we thank Fred Dick for this suggestion).

2012; Over, 2009; and recent special issue of *Thinking and Reasoning*) is explicitly probabilistic. However, different theorists who fall under the "new paradigm," while in broad agreement, diverge on how probabilities figure in the psychology of human reasoning, For example, Pfeiffer and Kleiter (2009, 2010; see also, Pfeifer, 2013) adopt a mental probability logic approach based on deduction and an interval based probabilistic semantics. We, on the other hand, regard causal Bayes nets (CBNs; Pearl, 1988, 2000, 2001) as an account of the mental representations constructed in WM, i.e., in System 2, as interpretations of conditional sentences (see also, Sloman, 2005; Sloman & Lagnado, 2005). In this section, this is the approach we will contrast with LP with respect to two inferences which we argue can be handled by the probabilistic approach but not by LP. These inferences are (i) learning about the strength of the relation expressed in a conditional sentence when confronted with inconsistency and (ii) explaining away alternative causes.

Despite Stenning and van Lambalgen (2005) themselves proposing that suppression effects might be dealt with by causal Bayes nets, actual attempts to model these inferences using CBNs have only just begun (Fernbach & Erb, 2013). Sloman and Lagnado (2005) and Ali, Chater, and Oaksford (2011) both looked at conditional inference and CBNs but not explicitly at the classical suppression effects (but see, Oaksford & Chater, 2013). The idea is that conditionals describe dependencies which are represented as directed edges in a Bayes net.[4] This view commits one to more than just probability theory, e.g., CBNs assume the acyclicity of dependencies, directedness, faithfulness, and the parental Markov property. All these assumptions are about making inference more tractable but some of these assumptions have been questioned (for a review, see Rottman & Hastie, 2013). However, the potential of CBNs to model conditional reasoning has not been fully explored and so it would be premature to dismiss them solely on these grounds (Oaksford & Chater, 2013; Rottman &

[4] There are exceptions described in Oaksford and Chater (2013) but these are usually dismissed in the philosophical literature as not requiring the same analysis as real conditionals.

Hastie, 2013). The idea is that conditionals build the appropriate dependency structures in System 2, i.e., in WM. Parameterising these structural dependencies is achieved via conversational pragmatics, i.e., the speech act of asserting a conditional indicates the conditional probability is high, and prior knowledge in System 1, i.e., in LTM. Figure 1 shows a Bayes net with dependencies representing the conditionals C_p, C_s, and C_r. We begin by looking at how these representations can account for explaining away inconsistency and the suppression effect for MP. In doing so, we also address a few conceptual issues about the interpretation of disablers in CBNs.

FIGURE 1 ABOUT HERE

Inconsistency and Suppression. So when we are told that C_p: *If she has an essay to write (p) she will study late in the library (q)* and discover that p but $\neg q$, what do we do? An obvious answer, following the logic of the example of Uranus in celestial mechanics, is to hypothesize that there is an auxiliary assumption operative, i.e., C_r: *If the library is open (r) she will study late in the library (q)*, which is considered to be necessary but not sufficient for q, i.e., $\Pr(\neg q|\neg r)$ is high but $\Pr(q|r)$ is not, and the library is closed ($\neg r$). Given such a parameterization of Figure 1, while $\Pr(q|p)$ is high $\Pr(q|p,\neg r)$ is low. That is, assuming the library is closed explains away the apparent counterexample. The same set up also explains the suppression effect for MP when participants are given C_p, C_r and p as premises compared to when they are given just C_p and p. Being told about C_r leads them to consider whether she will study late in the library (q) when the library is closed ($\neg r$) and she has an essay to write (p), i.e., to evaluate $\Pr(q|p,\neg r)$. Thus suppression effects for MP will arise when disablers are explicitly represented.

In Byrne's (1989) data, the probability of endorsing the MP inference for C_p in the presence of a defeater, C_r, was around 30%, i.e. significantly above zero and significantly below 50%. LP can only predict that in the absence of knowledge about whether the library is closed or open (r), you can infer nothing about whether she studies late (q), which suggests that at best an endorsement rate of 50% is predicted. It might be possible to improve LPs fit to these data by counting the number of distinct defeaters, $r_i \to ab_1$, for a particular rule. Presumably the more r_i that are available for a particular abnormality proposition, ab_1, the greater the level of suppression. However, Geiger and Oberauer (2007) have shown that the frequency of defeaters or alternative causes matters more in suppressing inferences than the range of different types of defeater or alternative cause (see also, Fernbach & Erb, 2013). This finding is consistent with the probabilistic approach but not LP.

We just concluded that to explain suppression effects requires the explicit representation of defeaters. However, this is contentious. The addition of the explicit edge for C_r in Figure 1 to *explain away* the inconsistency may not be necessary because in the CBN approach C_p is a *statistical* dependency, i.e., it is not inconsistent with the occurrence of a p and $\neg q$ counterexample. Indeed, the reason that $\Pr(q|p) < 1$ is because of the assumed existence of disablers like C_r. At least, this is the rationale behind the noisy OR representation of alternative causes (Fernbach, Darlow, & Sloman, 2010, 2011; Pearl, 1988). While alternative causes are explicitly represented as directed edges in a CBN, disablers are only represented implicitly as probabilities less than one.

This approach represents a substantive psychological claim about the nature of the representations underpinning human inference and action, one which we have argued is unlikely to be true (Oaksford & Chater, 2010, 2013). Moreover, recently Fernbach and Erb (2013) have proposed a CBN model of *modus ponens* in causal conditional reasoning where disablers are represented explicitly as in Figure 1. Moreover, a similar CBN representation has been proposed by Rottman and Hastie (2013) as a general approach to conditional inference. So there is nothing inherent to the CBN approach that precludes the explicit representation of disablers. Moreover, once disablers are explicitly represented, their probability, e.g., *the library is shut* (Pr(¬*r*)), will rise on learning that *she has an essay* (*p*) but she *doesn't study late in the library* (¬*q*), thus explaining away the apparent counterexample.

While in broad agreement with these approaches, we have questioned Fernbach and Erb's (2013) approach in which all disablers are represented, at least all that a reasoner knows about (Oaksford & Chater, 2013). We now briefly consider why. All disablers, or at least those known to a reasoner, are represented in the CBN in Fernbach and Erb (2013) and in Rottman and Hastie (2013) because $\Pr(q|p, \neg s)$, i.e., causal power,[5] is treated as equal to 1 minus the aggregate disabling probability. Consequently, if there are no disablers this probability is 1 and the cause will necessarily bring about its effect. This factor provides the impetus to consider disablers. In the representation of C_p , $\Pr(q|p) = 1$ and consequently a *p* and ¬*q* observation *is* inconsistent with C_p. Cummins (1995) implicit suppression paradigm provides good evidence that people do recruit and explicitly represent alternative causes *and* defeaters in causal conditional inference even when presented with only a single conditional premise, like C_p. However, we doubt that all known disablers are ever explicitly represented as it would seem to place far too great a burden on working memory. Rather we have suggested that they are made explicit as needed in the dynamically unfolding situations that require agents to draw inferences to achieve their goals (Oaksford & Chater, 2013, pp. 369-370).

Perhaps someone's goal is to find the girl referred to in C_p. They know C_p and are told by the girl's Mother that she has an essay to finish. They naturally infer that she's in the library probably without explicitly considering disablers. Considering explicit disablers probably only happens if they subsequently learn ¬*q*. So on their way to the library they may learn from a friend that, for example, the girl was seen on the other side of town to the library. In order to guide further action to achieve the goal of finding the girl, our reasoner

[5] That is, the probability of the effect given the cause in the absence of alternative causes (Cheng, 1997).

then recruits C_r to explain away the apparent contradiction which leads them to consider alternative locations where the girl may study when the library is closed. We would argue that all of this reasoning is probabilistic, i.e., C_p is still regarded as a statistical dependency. A reasoner's impetus too explicitly represent the disabler comes not from the need to explain away the inconsistency *per se* but from *trying to achieve their goal*. So attempting to explain away the p and $\neg q$ observation is more likely to achieve the goal of finding the girl, than simply conceding that C_p is statistical and so there is no inconsistency. We are currently exploring this more dynamic view of inference as belief revision (Oaksford & Chater 2013), as are others (Hartmann & Rafie-Rad, 2012; Douven & Romeijn, 2011).

Learning and inconsistency. The fact that C_p is statistical and has presumably been learnt from occasions when the girl has or does not have an essay to write and does or does not study late in the library, provides an approach to inconsistency that LP cannot provide. That is, one can learn from the occurrence of a p and $\neg q$ observation that $\Pr(q|p)$ is lower than one first thought. Of course, if uncertainty about C_p only derives from knowledge of disablers, as in Fernback and Erb (2013) then this makes no sense. However, it seems far more likely that we learn about statistical relations like C_p initially without explicit knowledge of disablers, as in standard causal learning scenarios using contingency tables (for a summary of such research, see, Hattori & Oaksford, 2007). This knowledge is then further refined by uncovering disabling conditions. The latter process may not be obvious from disablers like C_r which derive from the way libraries work and so can be inferred from general knowledge. However, a disabler like *if Coronation Street is on, she does not study late in the library* is specific to the girl and would also have to be learnt. It seems that this more nuanced knowledge of her behavioural dispositions would have to be acquired later as it *qualifies* C_p.

If someone is ignorant of any disablers and cannot discover any in the current context, then the p and $\neg q$ counterexample can be treated as an observation and hence an opportunity to learn about $\Pr(q|p)$. Many of the conditionals in which people are interested, like C_p, express the habits, dispositions, intentions, and promises that underpin our folk psychological understanding of each other and which enable us to predict and coordinate our behaviours in the social world. Take one of the current author's (MO) disposition to buy his morning coffee at Pret a Manger on Euston Road whenever he goes to work, i.e., *if Mike goes to work, he buys his coffee at Pret on Euston Road*. Now if he fails to buy his coffee at this coffee shop one morning on his way to work, it is plausible that a myriad of possible disablers could be listed to save him or an interested observer from contradiction. But ultimately unlike an engineered causal mechanism like a car most of these defeaters are opaque not just to an observer but to MO.[6] In such a case, an observer must simply learn that MO's disposition to buy his morning coffee at Pret on Euston Road is less reliable than she first thought. Of course, the converse is also true, successfully predicting that MO stopped for coffee at this location today should be an occasion to learn that the disposition is more reliable.

Oaksford and Chater (2013) discuss the problems of dynamic inference where new information may alter the original probability distribution. Specifically, they consider a learning approach to explain the empirical data on the *modus tollens* inference (MT), i.e., with C_p as the conditional premise, a reasoner is given the new information that *she did not study late in the library* and infers *she did not have any essay to finish*. However, in the above example of looking for the girl the context is one in which it is known she had an essay to finish and what the person looking for her now learns is the classic inconsistency that p, $\neg q$, *and* $p \rightarrow q$ (see, Oaksford & Chater, 2013). Of course, this can be resolved by recourse to a

[6] We remain neutral on whether subjective uncertainty arises here as a consequence of ignorance of the full range of defeaters (e.g., chemical imbalances in MO that drove him to avoid caffeine that morning), or irremediable objective uncertainty in the world. Whatever the reason, people's cognitive system represents and draws inferences about degrees of belief that, largely, obey the probability calculus.

disabler. But one may just alter one's degree of belief in the conditional premise.[7] This strategy implies that the original conditional probability, $\Pr_0(q|p)$, is changed to a new conditional probability, $\Pr_1(q|p)$, which violates the invariance assumption for Bayesian conditionalization (Jeffrey, 2004). Oaksford and Chater (2013) discuss the ramifications of this violation at length but for now we just note how the new $\Pr_1(q|p)$ may be learnt.

Oaksford and Chater (2013) showed how one can learn from the experience of the inconsistency by using Bayesian learning to alter the degree of belief in C_p. This involves two models represented as Bayes nets, one representing a dependence model and one representing an independence model, familiar from Oaksford and Chater's (1994) optimal data selection model. The counterexample is more probable in the independence model. Consequently, by one iteration of Bayesian learning the probability associated with this conditional, $\Pr_0(q|p)$, which is also taken to be the probability of the dependence model, can be revised in a coherent way to a new lower value. Importantly, Oaksford and Chater (2013) show that revising the conditional probability in this way can provide much better model fits to the canonical data on abstract conditional inference tasks (Schroyens & Schaeken, 2003). In such tasks, using abstract material, it is implausible to hypothesize the people have access to disablers.

These learning effects are mediated by System 1, which is responsible for acquiring the dependencies and their associated strengths and so provide the building blocks of the CBN representations people construct in WM, i.e., in System 2. This approach to inconsistency is not available to LP which does not deal with how degrees of belief may be updated.

Explaining away alternative causes. Ali, Chater, and Oaksford (2011) demonstrated discounting effects for cases where a pair of conditionals describes convergent

[7] Oaksford & Chater (2013) point out that this strategy implies that participants respond to the MT inference counterfactually, i.e., with how likely they would have been to infer she did not to have an essay to write on learning she was not studying in the library given what they now know.

causes for the same effect for which LP provides no explanation. In their experiments, they presented participants with pairs of conditionals like C_p and C_s above, yielding two conditionals, $p \rightarrow q$ and $s \rightarrow q$, where p and s are alternative causes for the effect q. According to a probabilistic analysis (Morris & Larrick, 1995), *discounting* should be observed. So, after learning that the effect has occurred, *she is studying late in the library* (q), learning that one cause has occurred, e.g., *she had an essay to write* (p), should lead to reductions in one's degree of belief that the other cause (s), *she has a text book to read*, has occurred. In Ali et al's (2011) experiments this is exactly what they observed. Participants' degree of belief in s on being told that q and p had occurred was much lower than when they were only told that q had occurred. The discounting inference, or explaining away, is a very important novel contribution of the Bayesian approach (Chater, Goodman, Griffiths, Kemp, Oaksford, & Tenenbaum, 2011).

This inference pattern cannot be captured by LP. Given two alternative conditionals like C_p and C_s the minimal model that results from composition is simply the two conditionals, $p \rightarrow q$ and $s \rightarrow q$. Learning q, as in an affirming the consequent inference (AC), just licences the disjunctive conclusion, $p \vee s$, in which case nothing can be concluded individually about p and s, which is why this case leads to fewer endorsements of the AC inference (Byrne, 1989). Similarly, learning that s is true does not discount the possibility that p is true. $p \vee s$ and $p \wedge s$ are logically consistent. So still nothing can be concluded about p. Learning s could only lead you to believe $\neg p$ if $p \vee s$ were treated as exclusive-or but this pair of conditionals, i.e., the premises, does not logically rule out the possibility that she has an exam tomorrow and she has a textbook to read. Another possibility is that degrees of belief in these cases are being calculated as logical probabilities where the possibilities a connective does not rule out as false are treated as equiprobable. This is the approach adopted by the theory of extensional probabilities in mental models (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999), which Ali et al (2011) explicitly ruled out as a general account of discounting and augmentation inferences using pairs of conditionals like these (see also, Fernbach & Erb, 2013). In summary, LP would appear unable to account for discounting alternative causes in causal conditional inference when two causes converge on an effect.

Summary. A probabilistic approach, in which conditionals are represented as dependencies in causal Bayes nets can account for suppression effects (see also, Fernbach and Erb, 2013), for learning that the conditional probability has changed (Oaksford & Chater, 2007, 2013)), and for explaining away alternative causes (Ali et al, 2011). LP can explain neither of the latter two observations in belief updating, nor can it explain the graded effects observed in the explicit (Byrne, 1989) and in the implicit (Cummins, 1995) suppression paradigm.

However, conceptually the LP and probabilistic accounts of non-monotonic reasoning are not as distinct as these arguments portray (Pearl, 1988; Oaksford & Chater, 2007, pp. 115-118). Psychologically they both rely on accessing limited amounts of relevant information about defeaters and alternative causes from long term memory, i.e., System 1, and building a small scale model as the interpretation of the premises in System 2. Stenning and van Lambalgen (2005) describe the process of constructing both a minimal model interpretation in LP and a CBN interpretation *as reasoning to an interpretation*. It is important to both accounts that once that interpretation is reached, no further information is taken into account, i.e., the world of the model is closed (anything not explicitly represented is assumed not to be the case in the situation being modelled). The inferences that these interpretations licence only follow on this assumption. In particular, this assumption underpins explaining away in Bayesian accounts (Morris & Larrick, 1995). The idea that we reason only over a small scale model that provides an interpretation of the premises is of course completely familiar from mental models theory (Johnson-Laird, 1983, 2006). In the next and final section, we offer some further speculations on how this idea may be part of the solution to the problem of global computation.

Local and Global Computation

The goal of this paper has been to establish the credibility of single function dual process theory as a psychological account of non-monotonicity in human reasoning and to argue that it is a superior to attempts to apply logic programming approaches in artificial reasoning to these data. Both LP and the probabilistic approaches, involve constructing small scale models of conditional premises in System 2 and both require some global computation in System 1, to decide on the most plausible disablers or alternative causes to include in a model. The LP approach is to use abnormality propositions in order to maintain the consistency of System 2 and to render tractable the search for possible disablers in System 1. However, it does so at the cost of psychological reality, i.e., how is what is normal and abnormal learned? With respect to System 1, the probabilistic approach may not fare any better. Keeping System 1 probabilistically coherent would seem to involve maintaining a globally consistent joint probability table for all the propositions in LTM. But this would involve computations every bit as intractable as Reiter's M-operator. Consequently we appear to be on the horns of a dilemma. We have argued that we require a probabilistic approach to explain all the forms of belief revision that people engage in and to explain the empirical results but this leaves us no better off in explaining how we do this against the background of the Quinean and isotropic nature of human cognition.

While we do not suggest there are any easy solutions to these problems, we do think that consideration of some recent philosophy of science (Cartwright, 1983, 1999; Hacking 1983) may cast a different light on the nature of the problem. We first reconsider the properties that Fodor (1983, 2001) introduced in arguing that there was no theory of central cognitive processes and hence of System 1.

Quine and Isotropy

In the section we look at the both these Fodorian properties in turn, starting with the Quinean property. Why are central processes of belief fixation Quinean? Fodor argues that there are various properties of these processes that can only be interpreted as implicating the whole of our belief system. For example, the *simplest* revision is presumably the minimal one that would cause the least changes in our overall system of beliefs. Moreover, the most *plausible* defeater is again presumably the most likely one given everything we know. These holistic properties are directly related to probabilistic versions of the Ramsey test in which subjective conditional probabilities are determined by adding the antecedent to our stock of beliefs and reading off the resulting probability of the consequent. This cognitive process involves accommodating the antecedent by making minimal change to our existing beliefs.

Isotropy arises from the idea that in explaining a phenomenon or in working out how to solve a problem everything we know is potentially relevant. Thus in explaining why the car did not start when the key was turned we should not, for example, isolate our knowledge of cars from celestial mechanics. After all it remains possible that the car didn't start after the key was turned because a meteor smashed through the engine block. Our knowledge, according to Fodor, cannot be organised in such a way that such long-distance dependencies are impossible.

We now argue that these properties emanate from a philosophy of mind that closely tracks a particular view in the philosophy of science which has been rejected by philosophers like Cartwright (1983, 1999) and Hacking (1983). After Cartwright, we refer to the view of central processes that may emerge, as the *dappled mind hypothesis* (Cartwright, 1999).[8]

The dappled mind hypothesis.

Fodor (1983) is using an analogy between scientific inference and everyday inference in describing the central cognitive system, i.e., the nature of System 1 processing. Prima facie it seems that the notions of simplicity and plausibility people need can only be computed over the whole of world knowledge and this will require a globally consistent System 1 that can deliver to System 2 the most plausible information it needs to address its current inferential goals. On this account, our world knowledge is construed like a globally consistent true scientific theory. Cartwright (1999) refers to such an account as the *fundamentalist* position, i.e., there is one coherent set of laws that, if known, would describe the whole world accurately. Cartwright (1983, 1999) argues that this position is untenable. In particular she argues that it is our specific models, which guide actions in the world, like building a laser, that are the real candidates for truth and not theories. Most of the fundamental laws of physics only apply *all other things being right*. That is, just like starting a car, turning the key only works assuming the conditions are right and that no disablers are present. She also argues that overarching theories, for example quantum theory, are not strictly true of the world. Theory only really contacts the world via specific models that set various parameters to certain values and ensures that various conditions are right. In this sense, models are the prime candidate for truth not theories. Moreover, our models are frequently inconsistent with each other. There are apparently, for example, several inconsistent models of the laser all of which find application in predicting the behaviour of particular devices.

[8] There is a certain irony in appealing to Cartwright's philosophy of science in the context of the arguments put forward in this paper, as she is one of the principal detractors of the Bayes net approach to causation.

Cartwright's dappled world hypothesis is that while some parts of the world behave in a law like manner others may not. Ian Hacking (1983, p. 219) put it as follows:

> "God did not write a Book of Nature of the sort that the old Europeans imagined. He wrote a Borgesian library, each book of which is as brief as possible, yet each book of which is inconsistent with every other. No book is redundant. For every book there is some humanly accessible bit of Nature such that that book, and no other, makes possible the comprehension, prediction and influencing of what is going on."

A slightly weaker more epistemological version of this hypothesis is that we can only ever construct local models of bits of the world in order to predict and explain what is going on in our immediate concrete context but can never hope to have an overarching consistent theory. Our proposal is to take this epistemological view as an account of the central cognitive system. That is, whether the world is dappled or not, we have a dappled mind.

The dappled mind hypothesis suggests that the limited models we construct in System 2 to guide our actions in the world take precedence, i.e., we are concerned in each context in which we must act that they are as accurate as they can be to allow successful prediction and action. So, inconsistencies must be repaired in our models by adding disablers or learning as we have outlined. But our need to draw inferences about the world is generally context bound. Models that work in one context may not work in another. The contents of System 1, from which people build their local models, do not form an overarching theory each part of which is consistent with every other part. That is, there is no imperative for the whole of a cognitive agent's world knowledge to hang together as a consistent whole. Given our inability to do this for scientific theories it seems a big ask to expect this of the cognitive system of individual agents. What is important is using whatever knowledge there is to hand to build a model in System 2 that can be repaired to more accurately predict what is going on in our immediate context if things go wrong.

However, constructing such models may not require computing intractable Quinean and isotropic properties in System 1. The problems of bringing to mind relevant information seems to underestimate the importance of the immediate concrete or deictic context in which people must act and the fact that inferential behaviour is usually goal directed, as we argued above. Most often, simply where we are provides a rich set of cues that bring to mind the relevant information we need. Moreover, our goals in a context similarly cue the information we need for their attainment. Indeed our goals are crucial for directing attention to the relevant information within our deictic context which provides the cognitive context of most of our inferential behaviour. That is, in most human reasoning the contents of our models in System 2 are a function of what we need to attend to in our immediate context in order to achieve our goals. Moreover, contra isotropy, most often people's models are extremely shallow and it is only our social embedding that allows us to transcend them not the isotropic nature of the cognitive system.

For example, assume MO is in his study at home with the goal of getting to work, which will involve using the car. His sub goal is to start the car, for which attending to its colour or many other of its properties, is initially at least irrelevant. The physical presence of the car, the key in his hand etc. are all concrete cues accessing information in System 1. When he gets into his car he will turn the key without any consideration of possible defeaters or alternative causes. Only if this action fails to produce the desired effect, the car does not start, will he consider possible defeaters. These will already be prepotent in System 1 triggered by the deictic context he is in. If he considers, for example, that the battery is flat, this will lead him to consider alternative causes, like bump starting. Note that the defeater must come first as this determines his choice of alternative cause, hot wiring will not work with a flat battery. If bump starting does not work then he will probably be stumped, i.e., this one level in the default hierarchy is the limit of his knowledge. Rather than stand around attempting an impossible isotropic inference, he simply rings the AA (Automobile Association, a UK based Roadside assistance scheme) to come and fix it. The fact that distant pieces of knowledge may be relevant does not mean it makes sense for him to attempt to establish their relevance. It may be that the reason his car is not starting is because of some cosmic ray induced quantum effect in the complex computer system controlling the engine of modern cars. Even if he had the relevant knowledge to establish this connection it would not help to achieve his goal of getting to work. This goal is much more easily achieved by contacting the AA and not taxing his limited cognitive resources any further. The knowledge on which we rely to achieve our goals is not all inside our own heads, it is socially distributed, and we use this fact to avoid unnecessary cognitive effort.

In summary, this section has been very brief and highly speculative and we have introduced a range of issues that require a much more detailed treatment. However, we felt it necessary to offer at least a glimpse of how we view the nature of System 1 and how it interacts with System 2 that addresses, however superficially, the problems identified for central processes by Fodor (1983). This is because Fodor (1983) was the point of departure for our original critique of logicist cognitive science (Oaksford & Chater, 1991). As has been argued in the philosophy of science, our models of the world parameterized to concrete situations may be the primary candidates for truth and hence provide the principle guides to successful action. These are constructed in System 2 and flexibly adjusted to match the unfolding events about which we need to draw inferences by recruiting information from System 1. While the current model in System 2 is consistent with the active parts of System 1 it seems unlikely that System 1 is a globally consistent system. Moreover, it seems unlikely that the epistemological problems created by the putative Quinean and isotropic nature of human reasoning are problematic for System 1. Properties like plausibility and relevance are most likely only quite local computations based on the quite shallow knowledge of complex systems most people possess. This is not generally problematic for achieving our goals given the social distribution of knowledge. Consequently, it may be that fixes to the technical frame problem in AI like those offered by LP, need not be part of an eventual theory of human reasoning.

Conclusion

In conclusion, probabilistic single function dual process theory provides a better psychological account of non-monotonicity in human reasoning than recent attempts to apply logic programming approaches in artificial intelligence to these data. LP only provides an approach that avoids global processing in the one pscyhological task to which it has been applied when all information about defeaters and alternative causes is explicit. To account for implicit suppression tasks and to distinguish disablers from alternative causes requires access to word knowledge. LP also is unlikely to be psychologically real, offering a solution to the technical frame problem only if it assumed that all we have to do is index conditions as abnormal. Like hash coding, LP provides no account of how these indices are learned. Moreover, there are a range of inferences concerning conditionals to which LP cannot be applied but for which a probabilistic approach using CBNs accounts naturally. However, there are a range of problems about the nature of System 1 and its relation to System 2, introduced by Fodor (1983), which remain problematic for both approaches. In a final speculative section, we suggested that these may be less problematic for the actual, context bound, and goal directed inferential behavior that guides our actions moment by moment in the everyday world. While many problems remain, the probabilistic single function dual processing approach remains the most promising account of the nonmonotonicity of human reasoning.

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford: CLSI Publications.

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional inference: Causal models or mental models. *Cognition*, *119*, 403-418. doi: 10.1016/j.cognition.2011.02.005.

Austin, J. L. (1962). *How to do Things with Words: The William James Lectures delivered at Harvard University in 1952*. J. O. Urmson and M. Sbisà (Eds.), Oxford: Clarendon Press.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 1-21. doi: 10.1016/0010-0277(89)90018-8.

Cartwright, N. (1983). *How the laws of physics lie.* Oxford: Oxford University Press.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*, Cambridge: Cambridge University Press.

Chater, N. & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture. *Cognition*, *34*, 93-107. doi: 10.1016/0010-0277(90)90033-G.

Chater, N., Goodman, N., Griffiths, T., Kemp, C., Oaksford, M., & Tenenbaum, J. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral & Brain Sciences*, *34*, 194-196. doi: 10.1017/S0140525X11000239.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367-405. doi:10.1037/0033-295X.104.2.367

Cummins, D. (1995). Naive theories and causal deduction. *Memory & Cognition*, *23*(5), 646-658. doi:10.3758/BF03197265

Cummins, D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274-282. doi:10.3758/BF03211151

Davidson, D. (1974). On the very idea of a conceptual scheme. In D. Davidson (Ed.), *Inquiries into truth and interpretation* (pp. 183-198). Oxford: Oxford University Press.

Douven, I., & Romeijn, J.-W. (2011). A new resolution of the Judy Benjamin problem. Mind, 120, 637–670. doi: 10.1093/mind/fzr051.

Evans, J. StB. T. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychological Bulletin*, *128*, 978–996. doi: 10.1037//0033-2909.128.6.978.

Evans, J. St.B. T. (2003). In two minds: Dual processes account of reasoning. *Trends in Cognitive Sciences, 7*, 454-459. doi: 10.1016/j.tics.2003.08.012.

Evans, J. St.B. T. (2007). *Hypothetical thinking*. Hove, East Sussex: Psychology Press.

Evans, J. St.B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.

Evans, J. St.B. T., & Stanovich, K. E. (2013). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on Psychological Science*, *8*, 263-271. doi:10.1177/1745691613483774.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, *21*, 329-336. doi:10.1177/0956797610361430.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal Of Experimental Psychology: General*, *140*, 168-185. doi:10.1037/a0022100.

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *39*, 1327-1343. doi:10.1037/a0031851.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press, Cambridge

Fodor, J. A (1983). *The modularity of mind*. Cambridge: MIT Press.

Fodor, J. A (2001). *The mind doesn't work that way*. Cambridge: MIT Press.

Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition, 35*, 2060-2074. doi: 10.3758/BF03192938.

Grégoire, E., Mazure, B. and Saïs, L. (1998). Logically-complete local search for propositional non-monotonic knowledge bases. In I. Niemelä and T. Schaub (Eds.), *Proc. 7th Int. Workshop on Nonmonotonic Reasoning* (pp. 37–45), Trento. Doi: 10.1.1.48.4725.

Haack, S. (1974). *Deviant logic*. Cambridge: Cambridge University Press.

Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for co-variation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, *31*, 765-814. doi: 10.1080/03640210701530755.

Hartmann, S., & Rafiee Rad, S. (2012). *Updating on conditionals = Kullback-Leibler distance + causal structure*. Presented at the Biennial Meeting of the Philosophy of Science Association, San Diego, CA, November 15–17.

Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *36*, 805-812. doi: 10.1037/a0018784.

Jeffrey, R. (2004). *Subjective probability: The real thing*. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2006). *How we reason*. Oxford: Oxford University Press.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88. doi: 10.1037/0033-295X.106.1.62.

Kahneman, D. (2011). *Thinking, fast and slow*. London, UK: Penguin Books.

Klauer, K. C. , Beller, S., & Hütter, M. (2010). Conditional reasoning in context: A dual-source model of probabilistic inference. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *36*, 298-323. doi: 10.1037/a0019445.

Kowalski, R. (2010). Reasoning with conditionals in artificial intelligence. In M. Oaksford, N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 254-282). New York, NY US: Oxford University Press.

Manktelow, K. I. (2012). *Thinking and reasoning*. Hove, East Sussex: Psychology Press.

McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, *28*, 89–116. doi:10.1016/0004-3702(86)90032-9.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence*, *3*, 151-160. doi: 10.1111/j.1467-8640.1987.tb00183.x.

Miller, G. A. (1996). Contexuality. In J. Oakhill and A. Garnham (Eds.), *Mental models in cognitive science* (pp. 1-18). Hove, East Sussex, UK: Psychology Press.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, *102*, 331–355. doi: 10.1037//0033-295X.102.2.331.

Oaksford, M. & Chater, N. (1991). Against logicist cognitive science. *Mind & Language*, *6*, 1-38. doi: 10.1111/j.1468-0017.1991.tb00173.x.

Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608-631. doi: 10.1037//0033-295X.101.4.608.

Oaksford, M. & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, Sussex: Psychology Press.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349-357. doi: 10.1016/S1364-6613(00)01699-5.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2009). Precis of "Bayesian rationality: The probabilistic approach to human reasoning." *Behavioral and Brain Sciences*, *32*, 69-120. doi: 10.1017/S0140525X09000284.

Oaksford, M., & Chater, N. (2010). Cognition and conditionals: An introduction. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals*: *Probability and logic in human thinking* (pp. 3-36). Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2011). Dual systems and dual processes but a single function. In K. I. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St.B. T. Evans* (pp. 339-351). Hove, Sussex: Psychology Press.

Oaksford, M., & Chater, N. (2012). Dual processes, probabilities, and cognitive architecture. *Mind and Society*, *11*, 15-26. doi: 10.1007/s11299-011-0096-3.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking and Reasoning*, *19*, 346-379. doi:10.1080/13546783.2013.808163

Over, D. E. (2009). New paradigm psychology of reasoning: Review of "Bayesian Rationality." *Thinking & Reasoning*, *15*, 431-438. doi:10.1080/13546780903266188

Over, D. E., Evans, J. StB. T., Elqayam S. (2010). Conditionals and non-constructive reasoning. In M. Oaksford and N. Chater (Eds.), *Cognition and conditionals:*

Probability and logic in human thinking (pp 135–152). Oxford: Oxford University Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Fransisco, CA: Morgan-Kaufmann, 1988

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Pearl, J. (2001). Bayesianism and causality, or, why I am only a half-Bayesian. In D. Corfield and J. Williamson (Eds.), *Foundations of Bayesianism* (pp. 19-36). Holland: Kluwer Academic Publishers.

Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logic perspective. *Thinking & Reasoning*, *19*, 329-345. doi:10.1080/13546783.2013.838189.

Pfeifer, N., & Kleiter, G. D. (2009). Mental probability logic. *Behavioral and Brain Sciences*, *32*, 98-99. doi: 10.1017/S0140525X09000442.

Pfeifer, N., & Kleiter, G. D. (2010). The conditional in mental probability logic. In M. Oaksford, N. Chater (Eds.) , *Cognition and conditionals: Probability and logic in human thinking* (pp. 153-173). New York, NY US: Oxford University Press.

Putnam, H. (1974). The ‘corroboration’ of theories. In A. Schilpp (Ed.), *The Philosophy of Karl Popper* (Vol. 2), La Salle, Ill.: Open Court.

Pylyshyn, Z. (1984). *Computation and cognition*. MIT Press, Cambridge

Reiter, R. (1985/1978). On reasoning by default, in R. Brachman and H. Levesque (Eds.), *Readings in knowledge representation*, Morgan Kaufman, Los Altos, CA, pp. 401–10.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*, 129-134. doi: 10.1111/1467-9280.00322.

Rips, L. J. (2002). Reasoning. In H. F. Pashler (Series Ed.) and D. L. Medin (Vol. Ed.), *Stevens' handbook of experimental psychology: Vol. 2. Cognition* (3rd ed.). New York: Wiley.

Rottman, B., & Hastie, R. (2013). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*, 109-139. doi:10.1037/a0031903

Rumelhart, D. E., McClelland, J. L., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I.* Cambridge, MA: MIT Press.

Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 140-149. doi: 10.1037/0278-7393.29.1.140.

Sellen, J., Oaksford, M., & Gray, N. (2005). Schizotypy and conditional inference. *Schizophrenia Bulletin*, *31*, 105-116. doi: 10.1093/schbul/sbi012.

Shanahan, M.P. (1997). *Solving the frame problem: A mathematical investigation of the common sense law of inertia*. Cambridge, MA: MIT Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22. doi:10.1037/0033-2909.119.1.3

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY, US: Oxford University Press.

Sloman, S.A., & Lagnado, D.A. (2005). Do we do? *Cognitive Science, 29,* 5-39. doi: 10.1207/s15516709cog2901_2.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford: Oxford University Press.

Stanovich, K.E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences. 23*, 645–726 . doi: 10.1017/S0140525X00003435

Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*, 919-960. doi:10.1207/s15516709cog0000_36

Wason, P. C, & Evans, J. StB. T. (1975). Dual processes in reasoning. *Cognition*, *3*, 141–154. doi: 10.1016/0010-0277(74)90017-1

Wernhard, C. (2011). *Forward human reasoning modelled by logic programming modelled by classical logic with circumscription and projection*. Technical Report Knowledge Representation and Reasoning 11-07, Technische Universität Dresden. Retrieved from http://www.wv.inf.tu-dresden.de/Publications/2011/report11-07.pdf.

Figures

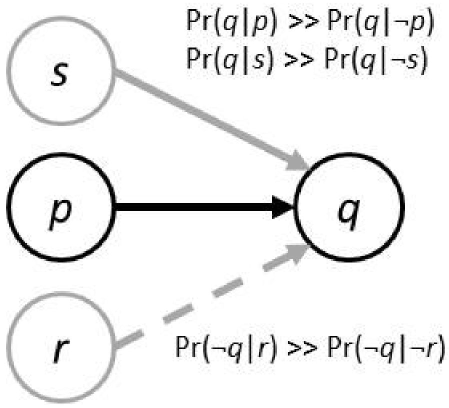

Figure 1. Bayes Net representations of the conditional premise (black) and associated alternatives (*s*) and additional (*r*) antecedents (grey). The priors Pr(*p*), Pr(*r*), and Pr(*s*) are also required in the parameterization.