

### **The psychological basis of the Harman-Vogel paradox**

One of the central puzzles of contemporary epistemology concerns the rule of single-premise closure. Here is the rule, as formulated by John Hawthorne:

If one knows P and competently deduces Q from P, thereby coming to believe Q, while retaining one's knowledge that P, one comes to know that Q. (2005, 29)

The single-premise closure principle (henceforth: closure) looks extremely plausible: competent deduction from what is known might well be considered the paradigm of a secure way to attain knowledge. Hawthorne considers it "unquestionably odd" to deny that a person knows that Q when he has satisfied the conditions of the antecedent (2005, 30) and similar sentiments are expressed by many other epistemologists. Keith DeRose claims that denying closure yields "intuitively bizarre" results (1995, 28); Sherrilyn Roush thinks that the majority is right to find rejecting closure "too hard to swallow intuitively, even for a nice answer to skepticism" (Roush, 2005, 41) and Timothy Williamson, who actually names a stronger multi-premise version of the principle "intuitive closure", urges that we should be "very reluctant to reject intuitive closure, for it is intuitive" (2000, 118).

Notwithstanding its intuitive appeal, closure might seem to be threatened by certain patterns of judgment which also seem intuitive to us. One such pattern was identified by Gilbert Harman in the early 1970s. As Harman points out, it seems natural to deny that a person can *know*, prior to the draw, that his ticket in a fair million-to-one lottery will not win. However, it can also seem natural to grant that a person can know certain propositions that he would know to entail that the ticket is not going to win. I might know that I am meeting a friend for lunch tomorrow, Harman contends, even if I also know that his being there would be incompatible with his having won the lottery – say, on the assumption that the winner would have to go upstate to collect the prize (Harman, 1973, 1986).

If Harman's pattern of intuitions presents a challenge to closure, it is not because this pattern on its own constitutes a violation of the principle. One can share Harman's initial intuitions about what is known without feeling that it would be right to try to deduce the proposition about the ticket not winning from proposition about the upcoming lunch, and to claim knowledge of the ticket's loss on that basis. What is not immediately clear is why that deduction feels wrong. One

possible reason has been highlighted by Jonathan Vogel in his defense of closure: if we try to deduce that the lottery ticket will not win from our expectation of the lunch meeting, we get the feeling that we no longer know the lunch proposition, or perhaps never did know it (Vogel, 1990). If attempting the deduction results in a loss of knowledge of one's initial premise, closure is not violated: retaining that knowledge throughout the deduction is one condition of the antecedent of the principle. The problem is that preservation of closure now seems costly: we seem to have lost our right to count ourselves as having knowledge of certain mundane facts we would once have counted ourselves as knowing.

Focusing on certain features of Harman's original example, it may seem that only a small range of our ordinary knowledge is in peril. It is sometimes suggested that Harman's pattern of intuitions should generate worries only about our knowledge of future events (Dretske, 2005), or of outcomes dependent on propositions whose support is purely statistical or probabilistic in character (Cohen, 1988; Nelkin, 2000). Unfortunately, similar patterns of intuitive assessment can be triggered with examples concerning present events and without overt reference to probabilities (Vogel, 1990). For example, considering an ordinary person who parked his car an hour ago on Avenue A and remembers having done so, Vogel notes we would ordinarily say this person – call him Smith – knows his car is on Avenue A.<sup>1</sup> Vogel then points out that when we consider whether Smith now knows that his car is *not* one of the hundreds of cars that are stolen each day across the country, it is hard to evaluate him so positively. One is naturally inclined to say that Smith does *not* know that his car is not one of the ones that has been stolen and driven away, even though this is a proposition which is entailed by the proposition we initially took him to know, and even though we would have no trouble in crediting Smith with knowledge of this entailment.

Although it may not be immediately evident what the lottery and car theft cases have in common, having read Harman and Vogel's examples, it is easy to generate more instances of this pattern. To use an example of Igor Douven's, I find it natural to say that I know that I will be teaching epistemology next fall, but much harder to say that I know that I will not be killed in a traffic accident over the summer (Douven, 2007). Yet on reflection, it is clear to me that my continued survival is one of the many conditions that will need to be met if I am to teach anything at all come September. Given my background beliefs, it would be incoherent for me to accept that I

---

<sup>1</sup> Vogel is on solid ground in claiming that we would use 'knows' here: 'know' is one of our most common verbs, used more often than 'thinks' and 'believes' put together (Leech, Rayson, & Wilson, 2001), and there is empirical evidence that it is normally used in cases like this one (Booth, Hall, Robison, & Kim, 1997; Frank & Hall, 1991).

am going to be teaching epistemology in the fall while refusing to accept that I am not going to be killed in a traffic accident over the summer.

There is something alarming about the ease with which we can generate further instances of the problematic pattern of intuitions. If the correct response to Harman's original example is to comply with the closure principle and deny that one has knowledge of the next day's lunch meeting, then it seems that very similar considerations should push one to deny that anyone has knowledge of a very great range of everyday matters (Christensen, 2004; Hawthorne, 2004). At this point one might be tempted to reconsider one's loyalty to the closure principle. Dretske urges us to reject it, on pain of skepticism: "For anyone who thinks they know something about the material world, the choice is easy. Closure must be denied." (Dretske, 2005, 23)

Many responses have been offered to the Harman-Vogel paradox in recent years,<sup>2</sup> most notably a series of innovative rival proposals about either the semantics of knowledge-ascribing sentences, or the metaphysical determinants of knowledge itself. Semantic proposals such as contextualism (Cohen, 1999; DeRose, 2009; Lewis, 1996) and relativism (MacFarlane, 2005; Richard, 2004) hold that the verb "to know" has some surprising features. Contextualists aim to explain the shifting intuitions of the paradox by claiming that "knows" picks out different relations in different conversational contexts. Once worries about remote possibilities have been raised, "knows" comes to denote a more demanding relation, so that a subject needs to be in a stronger epistemic position to count as knowing. Relativists take the more radical step of saying that the truth of a knowledge-ascribing sentence for an assessor depends in part on a hidden variable, the assessor's current epistemic standards parameter. Meanwhile, metaphysical proposals such as interest-relative invariantism argue that the paradox is better handled by taking knowledge to involve some factors overlooked in traditional epistemology, factors such as the subject's practical concerns (Fantl & McGrath, 2007; Hawthorne, 2004; Stanley, 2005). On such views, whether a subject knows a given proposition may depend in part on what he is worried about, or whether that proposition is a high-stakes matter for him. It is safe to say that each of these novel semantic and metaphysical theories is thought to have a number of counter-intuitive consequences.

---

<sup>2</sup> This pattern of intuitions is sometimes called the "Lottery Paradox", although as Igor Douven notes (2007), calling this puzzle about knowledge the Harman Paradox helps distinguish it from the distinct paradox of rational acceptance identified in (Kyburg, 1961). I also think that Vogel is right that reference to lotteries is not essential to the pattern, hence the broader name.

For those who resist such semantic and metaphysical innovation and instead adhere to a more traditional view of knowledge, “knows” always denotes the same relation, and the question of whether someone knows is entirely settled by factors such as the truth of the proposition believed and the quality of the subject’s thinking. Traditional or ‘intellectualist invariantist’ epistemologists seem to face an especially unpleasant challenge in the Harman-Vogel pattern of intuitions. Because the hard-to-know proposition is entailed by the one that seems easier to know, and because competent deduction is a top-quality cognitive process, it may seem that intellectualist invariantists must appeal to an error theory to explain the core intuitions of the paradox. For example, Douven has advanced an account in which a person who says she knows she will meet her friend for lunch the next day does not literally mean that, but really means instead that she knows she will meet her friend *unless something very unlikely occurs*, where this last proviso is for pragmatic reasons typically left unspoken (Douven, 2007). If it doesn’t seem obvious that we typically have such a proviso in mind, defenders of this approach will have to work hard to convince us otherwise.

Faced with the Harman-Vogel paradox, one might wonder whether there is any way of defending a traditional view of knowledge in which our common knowledge-ascribing sentences literally mean what they say. My aim in what follows is to provide such a defense. Instead of attempting any pragmatic, semantic or metaphysical innovation, I seek to explain the intuitions behind the paradox in terms of some very basic features of the psychology of human judgment. Drawing on a body of empirical work known as “dual-process theory,” I argue that the two apparently conflicting judgments in the pattern are naturally made in different modes of cognition. The easy call that one knows one will be teaching the course or meeting someone for lunch is made in an automatic or heuristic mode – known as ‘System 1’ in the psychological literature, after (Stanovich, 1999) – while the harder assessment about knowledge of lottery ticket loss or traffic accident survival must be made in a controlled or analytic mode (‘System 2’). This switch in modes explains why we face difficulties in deducing the harder claim from the easier one, despite our recognition of a relevant relation of entailment. Explicit deduction is invariably a controlled (System 2) process, and in the context of the Harman-Vogel pattern, shifting into System 2 tends to bar us from making the original automatic System 1 judgment, for reasons to be discussed in some detail in what follows. However, the fact that we are not always able to make the original System 1 judgment does not show that this judgment was in any way defective when initially made in that mode: I argue that an automatic judgment of a certain proposition may fully deserve to count as knowledge even if there are conditions under which that particular judgment could not be produced automatically.

Section 1 of the paper describes the Harman-Vogel pattern of intuitions in more detail, and discusses the shifts the pattern that appear under pressure, such as the inclination to rescind one's original judgment that one *knows* one is meeting one's friend for lunch tomorrow when it is pointed out that various contingencies might block that outcome. Section 2 of the paper explains some of the differences between the two modes of cognition posited by dual-process theory, and outlines some of the factors that trigger switches between one mode and the other. These switches are determined either by features of the subject matter being judged, or by contextual factors relating to the circumstances of judgment: Section 3 aims to identify the relevant features active in the Harman-Vogel pattern and its variants, including cases involving compound lotteries, which seem to give us intuitions contrary to those originally identified by Harman. Section 4 argues that the dual-process analysis allows us to retain the principle of single-premise closure for knowledge ascriptions. Finally, although appeal to mode-shifting might still seem hospitable to epistemological programs such as contextualism or interest-relative invariantism, Section 5 argues that there is nothing in the dual-process account to rule out traditional intellectualist invariantism about knowledge.

## 1. Identifying the Harman-Vogel pattern

In the Harman-Vogel pattern of intuitions, it initially seems intuitive to say that a subject knows a certain proposition P, and that another proposition Q is entailed by P, but counter-intuitive to say that the subject knows that Q. Some terminology may be useful in discussing this pattern. The hard-to-know proposition Q has been given several names in the literature: Dretske calls it a heavyweight proposition, Vogel calls it a lottery proposition (notwithstanding his own generalization of the problem to cases not involving lotteries). In what follows I'll use the label 'stringent' for the Q-side of the pattern, and 'lax' for the P-side. To distinguish the subject's cognitive act from the ascriber's, the first will be labeled the lax (or stringent) *judgment*, and the second the lax (or stringent) *ascription*; the underlying proposition judged – or in some cases, potentially judged – by the subject will be labeled the lax (or stringent) *proposition* on the basis of its initial position in the table.<sup>3</sup> Although the ultimate aim of this paper is to clarify what is

---

<sup>3</sup> To stave off a possible terminological confusion: lax propositions do not *always* receive lax evaluations. To identify a proposition like "I will be meeting my friend for lunch tomorrow" as a lax proposition is just to indicate that in normal circumstances it is the sort of thing that could seem easily known, for reasons to be

happening psychologically at the level of ascription, considerable attention will be paid to the level of judgment; indeed, it will be argued that certain shifts in the psychology of ascription are inherited from shifts occurring at the level of judgment. Given that the ascriber must either retrace or represent the steps of the subject in making her ascription, whether this is by simulating the subject's cognition, or by theorizing about how the subject would think, some features of judgment will carry over to the level of ascription. If a certain type of proposition can be grasped only with the help of controlled cognition, for example, then some degree of controlled cognition will be needed both for the judgment and for the ascription of this type of proposition.

The following chart lists a variety of examples of the Harman-Vogel pattern, taken from the work of Harman, Vogel, Dretske and Hawthorne:

<b>(a) Apparently easy P ('lax')</b>	<b>(b) Apparently harder Q ('stringent')</b>
i. Bill knows that he is meeting his friend Mary for lunch tomorrow.	i. Bill knows that Mary's ticket in tomorrow's fair one-in-a-million lottery will not win.
ii. Lucy knows that the current US President is Barack Obama.	ii. Lucy knows that no one has assassinated Barack Obama in the last hour.
iii. Looking at the striped animal in the zoo, Fred knows that it is a zebra.	iii. Looking at the striped animal in the zoo, Fred knows that it is not a cleverly disguised mule.
iv. Smith knows that his car is on Avenue A, where he parked it an hour ago.	iv. Smith knows that his car is not one of the hundreds of cars that are stolen every day across the country.
v. John knows that he will be spending part of the summer in Syracuse.	v. John knows that he is not 'one of the unlucky people who, despite being apparently healthy, suffer a fatal heart attack in the next week.'
vi. Joe, a person of modest means, knows that he will not have enough money to go on an African safari this year.	vi. Joe, a person of modest means, knows that he will not win a major prize in a lottery this year.

For each example pair listed here, the reader needs to make several assumptions to get the standard initial intuitions. To generate a sense of paradox, it is important throughout that one conceive of the subject in the left-hand column as having the same information about the world as his or her right-hand counterpart. The subject must be assumed to have enough evidence both to

---

discussed in Section 3 below. Section 3 also discusses the special circumstances in which it is difficult to claim knowledge of such lax propositions.

make the judgment in the lax column seem natural, and to see that the stringent proposition is entailed by the lax one (perhaps with the help of some background information whose status as knowledge is not in question). The subject's evidence is not however taken to be of a sort that would make it immediately intuitive to ascribe knowledge of the stringent proposition to him or her. As was clear in the original context of these examples, we are supposed to flesh out these scenarios so that, for example, Lucy is an average citizen who last saw Obama on TV yesterday, rather than one of Obama's bodyguards currently on duty; Fred is looking at the striped animal through the fence 'at some distance' and without binoculars; Smith remembers parking his car on Avenue A but does not presently have a clear view of his parking spot, and so on.

There is some controversy over exactly which pairs of judgments deserve a place in the list: Vogel, for example, argues that Dretske's case (iii) doesn't really belong, because zoo visitors looking at zebras do have background information on the normal workings of zoos that would enable them to know both the lax proposition and the stringent one. Perhaps the focus on what Fred is currently seeing works to distract us from this, because the background facts about zoos are not visually supplied at the moment of Fred's judgment, but if we think about the case carefully, Vogel suggests, we should say that Fred really *does* know that the animal is not a cleverly-disguised mule (Vogel, 1990).

Without taking a position on whether Vogel is right about what Fred actually knows, it may be worth resisting his suggestion that this example be set aside as we search for an explanation covering contrasts in intuition like the ones produced by cases (i) and (iv). Our initial impression, even on Vogel's telling, was that Fred seemed not to know that the animal was not a disguised mule; if careful reflection tells against that intuition, bringing it into line with the initial lax ascription, this shift towards coherence is conceivably an instance of the same phenomenon we see when reflection is brought to bear on other pairs in the list. To be sure, coherence is often achieved by a shift in the lax rather than the stringent column: once we start thinking about the possibility of recent assassination, we may start to feel inclined to deny that any ordinary citizen knows who the President is at this instant. Claims to know what one is going to be doing next fall, or even at lunchtime tomorrow, can also be thrown into question once various tricky prospects are contemplated. If we engage in deliberate reflection on such contingencies as we look down the rest of the left-hand column, it may seem more intuitive to say that the subject in each case would be right to judge that the lax proposition is highly probable given his or her evidence, rather than that it is known. But unless there is some special reason to think that different forces are at work when

coherence is achieved by shifting one judgment rather than the other,<sup>4</sup> it is worth attempting the more ambitious project of developing a theory that will explain both our initial impressions across all the examples listed, and the shifts towards coherence that may emerge in either column on second thought.

Recognizing that one may become stringent about one's initially lax judgments does not necessarily dissolve the difference between these cautious second impressions and the judgments that were naturally stringent from the start, a point Hawthorne makes in a discussion of heavyweight judgments such as the judgment that one is not presently a brain in a vat. Hawthorne writes:

I have an immediate and strong inclination to say that I do not know such things in a way that I do not, for example, have any such immediate inclination to say that I do not know that I am working on a computer right now. Of course, I might derivatively generate some inclination to say that I do not know the latter via reflection on the former. But in the case of the former, doubts are natural and immediate, whereas in the case of the latter, doubts tend to be derivative. We should all at least agree that the relevant contrast is a salient fact about human psychology and that it is one of the bases of epistemic puzzlement. (2005, 33)

In what follows I will take seriously the suggestion that the contrast here is a fact about human psychology, and attempt to give it a psychological explanation.

## 2. Two modes of cognition

A number of separate research programs in cognitive and social psychology have in recent years converged on one major idea: there is a deep split between two kinds of human cognition, a split which is known to run through such epistemologically significant domains as probability

---

<sup>4</sup> Vogel's own diagnosis of the distinctive features of cases like (i) and (iv) is that these cases involve possible outcomes that are presented as statistically unlikely but "not ... *abnormal*, in some intuitive sense" (1990, 16). Vogel suggests there would be something *peculiar* about finding that the animal in the pen was a disguised zebra; to find that one's lottery ticket had won might be equally surprising but would not be equally strange. Indeed, Vogel points out, in the lottery and car theft cases we have some positive statistical evidence that gives some slight support to the claim that one's ticket has won or one's car has been stolen: in his estimation our background knowledge about auto theft rates works like our background knowledge about the odds of winning the lottery. Although I think there is something right about Vogel's focus on abnormality here, it may not be the only factor at work, and in any event would need to be spelled out with more precision to extend the account to other cases. Should a very recent assassination be considered abnormal, like a disguised zebra, or statistically slightly likely, given background Presidential assassination rates? Section 3 below offers an alternative general theory of the reasons – including reasons having to do with abnormality – for correcting in one direction rather than the other.



assessment, categorization and social judgment (Chaiken & Trope, 1999; Evans, 2007a; Frankish & Evans, 2009; Gilbert, 1989; Satpute & Lieberman, 2006; Sloman, 1996). Partly because the research programs identifying these two types of cognition originally developed somewhat independently of one another, they adopted different terminology to pick out the two sides of the contrast: a recent survey by Evans (Evans, 2007a) identifies fourteen different pairs of labels that have been used here, from Evans's own "heuristic vs. analytic" to Sloman's "associative vs. rule-based" (Sloman, 1996) to the "automatic vs. controlled" labels favored in (Schiffrin & Schneider, 1977).

The existing characterizations of the dichotomy differ not only in their terminology, but also on substantive points of detail and emphasis. For the purposes of neutral discussion, however, Keith Stanovich has proposed the adoption of the generic labels "System 1" and "System 2" to identify the underlying forms of cognition that the various formulations of the contrast have been attempting to capture (Stanovich, 1999).<sup>5</sup> In what follows dual-process theory will be discussed with the help of these generic labels, in a manner that draws only on the common core of generally recognized characteristics of each type of processing, and not on any controversial feature posited by some dual-process theorists but not others.<sup>6</sup> An exception to this neutral approach will be made near the end of section 3, where it will be expedient to discuss a specific model of dual-process theory – Evans's 'satisficing' model – in connection with a particular problem in knowledge ascription.

Known by the generic name 'System 1', the heuristic or automatic cognitive mode operates swiftly and without effort. This mode of processing is typically called into play to deal with familiar,

---

<sup>5</sup>Even these generic labels need to be handled with some care: Stanovich himself cautions against the assumption that 'System 1' should be understood as a singular entity, and in his own work often prefers the label TASS – The Autonomous Set of Systems – to underscore the heterogeneity of the automatic processes involved. In recent work he has also argued for a subdivision within 'System 2' between an algorithmic and a reflective level of processing (Stanovich, 2009). These finer-grained distinctions will not be called into play in what follows; the argument of the present paper depends only on the cruder and almost universally accepted System 1/System 2 distinction.

<sup>6</sup>It should be acknowledged that dual-process theories are very broadly but not universally accepted among psychologists: there are several nonstandard theories, including Kruglanski and Thompson's (1999) "unimodel" and the dynamic graded continuum model developed by Cleeremans and Jiménez in their (2002) and defended as an alternative to dual-process theories by Osman in her (2004). However, as pointed out by Adam Alter and colleagues, even these nonstandard theories not only predict transitions from simpler or more implicit to more complex or explicit forms of processing, but do so in the same domains where dual-process theories envisage a switch from System 1 to System 2 (Alter, Oppenheimer, Epley, & Eyre, 2007). Consequently, it should be open to advocates of such theories to adapt the explanation of the Harman-Vogel pattern given here in the terms of the dominant dual process theory. To my knowledge the only other active research program at odds with dual process theory is the Bayesian framework of (Oaksford & Chater, 2007); for present purposes I am setting aside the question of how (or indeed whether) the Harman-Vogel pattern could be explained within that particular model.

low-stakes and routine cases. System 1 enables us to respond immediately on the basis of prior knowledge or belief, often on the basis of associatively computed similarity to a mental model or stereotype. Because of its speed, and because it does not tie up working memory, System 1 serves as the default mode of judgment (Kahneman & Frederick, 2005). When we need to reason about novel or complex problems, however, we typically engage a more controlled mode of cognition. System 2 processing is serial in character: the output of one stage of cognition can be used as the input to the next, often with conscious awareness of successive stages of reasoning. System 2 is slower and more effortful than System 1, in part because it taxes our limited supply of working memory. It is called into play to handle unusual, unfamiliar or high-stakes problems, exceptions, most operations of negation, hypothetical and counterfactual reasoning (Evans, 2007a; Kahneman & Frederick, 2005; Sloman, 1996; Strack & Deutsch, 2004).

Simple arithmetical problems are sometimes used to illustrate the difference between the two systems. For example, Keith Frankish notes that when asked to multiply 9 by 11, most adults find that an answer comes to mind immediately and without effort. The System 1 processing here is considered autonomous: it generates an answer to the problem at once, whether or not the subject wants it to, on the basis of prior knowledge. The less common problem of dividing 9 by 11, on the other hand, is not answered automatically by typical adults. Unless one is a mathematical prodigy or has just memorized the answer to this particular problem, one will need to engage in the sequential reasoning of System 2 to find the solution. This type of reasoning is controlled: one is conscious of a series of steps, each taking the output of earlier stages as input, where each of these steps is itself an automatic computation arising from System 1 (Frankish, 2009; Sloman, 1996).

The explicit application of learned procedures in long division is a good example of System 2 processing, but this mode of thinking can take a great variety of other forms as well. An example of Steve Sloman's illustrates a difference in the ways the two systems tend to present themselves to consciousness:

When a response is produced solely by the associative system, a person is conscious only of the result of the computation, not the process. Consider an anagram such as *involnutray* for which the correct answer likely pops to mind associatively (involuntary). The result is accessible, but the process is not. In contrast, a person is aware of both the result and the process in a rule-based computation. Consider a harder anagram such as *uersoippv*. If you figured out the answer (purposive), and even if you did not, you likely applied various rules (like put different letters in the first position) which were consciously accessible. If you did get the answer, you were aware not only of result the but also of the process. (Sloman, 1996, 6)

A controlled effort to solve the harder anagram does not have to involve explicit internal verbal reasoning of the type involved in long division, but will still require working memory to hold various transpositions of letters in mind, to be used as input for further transformation. Although we are conscious of the series of steps involved in solving the harder anagram, and so in this case have excellent introspective evidence that our processing here is controlled, introspection is not an infallible guide as to which type of processing is engaged in solving any given problem.<sup>7</sup> There is evidence that we can be unaware of the fact that we have shifted between automatic and controlled cognition (Cary & Reder, 2002). Serial processing is much slower than associative processing, but can still be rapid enough to make clear introspective access difficult, and, as noted by Hume, deliberate introspection into the nature of one's current thinking may alter the character of what it investigates (Hume, 1739/1978, xix). Consequently, psychologists have in recent years turned to manipulations of working memory in order to determine which type of processing is engaged in various tasks. Given that involvement of working memory is taken to be the central characteristic of System 2 processing (Evans, 2007a), and given that we all have sharp limits on working memory space, performance on tasks requiring System 2 processing will suffer disproportionately under conditions which reduce available working memory, conditions such as distraction by other tasks (De Neys, 2006).

In the arithmetical and anagram examples the easy problem naturally triggers System 1 and the harder problem triggers System 2, but the selection of a system is not always determined by the content of the problem one encounters. For example, raising the subject's perceived stakes tends to activate System 2, even for problems which would ordinarily be handled by System 1 (Lerner & Tetlock, 1999). One might, say, run a systematic double-check of one's automatic answer to the first anagram if promised a large prize for getting it right.<sup>8</sup> Indeed, some of the most revealing cases for dual-process theory are problems for which either cognitive mode could be called into play, and for which the two modes will deliver different answers. Problems of this type figured in the emergence of one of the research programs that evolved into contemporary dual process theory – the heuristics and biases program of Tversky and Kahneman.

---

<sup>7</sup> Sloman himself is clear on this point: he notes that because both systems may be operative simultaneously, and because some controlled processing may not be entirely accessible to consciousness, conscious accessibility is a somewhat rough guide to which process is operative (1996, 6).

<sup>8</sup> Under high stakes we are inclined to think more in a more controlled fashion unless we are compromised by factors such as time pressure or restrictions in available working memory: subjects under immediate pressure to act and subjects distracted by another task may fail to shift up to System 2 even when stakes are high (Kruglanski & Webster, 1996; Lerner & Tetlock, 1999). For a discussion of the epistemological impact of this effect, see (Nagel, 2008).

The Harman-Vogel pattern concerns outcomes that seem easier to accept than certain higher-probability outcomes they entail, even by subjects who are capable of explicitly recognizing the entailment. Just this sort of pattern was found by Tversky and Kahneman in their examinations of our intuitive estimations of probability. Noting that our intuitive probability judgments do not always comply with the axioms of probability theory, they proposed that ordinary intuitive judgments are generated by a heuristic function which ranks outcomes as more or less probable on the basis of automatic assessments of their similarity to some mental model or stereotype. Tversky and Kahneman introduced the name 'representativeness' for the heuristic responsible for this intuitive method of assessing probability, suggesting that 'an outcome is representative of a model if the salient features match or if the model has a propensity to produce the outcome' (1983, 296). Because similarity to a stereotype can be enhanced by adding details to a description of an outcome, adding detail to a scenario can increase its apparent probability when we are judging on the basis of representativeness, even as the added details make the outcome objectively less probable.

In one well-known study, subjects read a description of a woman who is a strong match for the category 'active in the feminist movement'. The character, 'Linda,' was described as a 'single, outspoken, and very bright' 31-year-old who had been 'deeply concerned with issues of discrimination and social justice' as a student, and involved in anti-nuclear demonstrations (Tversky & Kahneman, 1983). Subjects were then presented with the three statements (a) Linda is active in the feminist movement, (b) Linda is a bank teller, and (c) Linda is a bank teller and is active in the feminist movement. These key statements were mixed in with five decoy sentences about other occupations (psychiatric social worker, elementary school teacher) and activities (yoga classes) that Linda might be pursuing. Subjects were asked to rank all eight statements in the list from the most to least probable. Almost 90% of them ranked (c) more probable than (b), violating the conjunction rule, according to which a conjunction can never be more probable than one of its conjuncts.<sup>9</sup> Graduate students in psychology, medicine and education did not fare better than naïve

---

<sup>9</sup> Worries about conversational implicature are often raised in response to these findings. Tversky and Kahneman themselves were concerned about this issue: perhaps subjects would read (b) in contrast with (c) and take it to describe an outcome in which Linda is a bank teller who is *not* active in the feminist movement. They themselves tested a version of the problem in which any potential conversational implicature was explicitly cancelled (1983, p. 299), and found that subjects continued to violate the conjunction principle. More recently, Stanovich and West have found in a between-subjects design that groups of subjects who are asked to evaluate only (b) relative to some decoys still give it a lower likelihood rating, on average, than the rating given to (c) by comparable subjects asked to evaluate only (c) relative to the same decoys. (Stanovich and West 2008).

first-year undergraduates in this task, and even PhD students in decision science who had taken advanced courses in statistics performed only marginally better, with 85% still violating the conjunction rule. When the decoy sentences were removed and subjects were simply asked to rank (a), (b) and (c), naïve students showed a slight improvement, and statistically sophisticated students showed a much greater improvement, with 64% now avoiding the fallacy. Interviews following the experiments confirmed that even naïve participants could recognize the legitimacy of the conjunction rule when this rule was presented explicitly: they simply failed to apply it in practice when presented with the concrete Linda problem. Most sophisticated subjects brought their judgments into compliance with the conjunction rule even without having the rule stated to them explicitly, but only when their attention was drawn to the applicability of the rule by the transparent (decoy-free) version of the problem.

To secure independent evidence that subjects who reach the right answer to the problem are employing a different mode of processing, researchers have examined differences between the performance of subjects who do and do not comply with the conjunction rule. For example, Wim De Neys has found that subjects who give the correct answer are in fact substantially slower in responding than subjects who judge that it is more likely that Linda is a feminist bank teller than that she is a bank teller. In addition, the percentage of people answering correctly drops drastically in pools of subjects who are burdened with a secondary task demanding working memory (De Neys, 2006).

Although there is some dispute over how exactly to characterize the heuristic processing that leads us into violation of the conjunction rule on problems like the Linda problem,<sup>10</sup> the consensus is that the natural prevalence of the rule violation is evidence that we do not ordinarily compute the probability of a complex outcome by compounding the various elementary probabilities upon which it depends (which would be very taxing, to say the least). There is evidence that even the subjects who give the correct answer will initially have the common automatic response, but recognize the need to override this response when they are reminded by

---

<sup>10</sup> For example, Vincenzo Crupi, Branden Fitelson and Katya Tentori have argued that what the mistaken subjects are doing is really substituting degree of confirmation for objective likelihood: because the “Linda” scenario yields more confirmation to hypothesis (c) than to hypothesis (b), the former seems more likely (Crupi, Fitelson, & Tentori, 2008). Crupi et al. are willing to grant that their proposal might actually be seen as a sharpening of Tversky and Kahneman’s loosely-worded heuristic, rather than as a rival to it (2008, 194). In any event this proposal fits with the broader dual-process framework: in subjects who get the problem wrong, a more accessible attribute (confirmation) is substituted for an attribute that is much harder to compute (probability).

the conjunction pattern – a pattern made much more noticeable when decoys are stripped away – that the automatic response pattern conflicts with a learned rule.<sup>11</sup> High-IQ individuals are more likely than lower-IQ individuals to override their automatic responses when cued to do so (Stanovich, 1999), but are not in general more likely to engage in systematic rather than automatic cognition in the absence of such cues (Stanovich & West, 2008).<sup>12</sup>

Problems for which System 1 and System 2 deliver different answers have had particular value in the development of dual-process theories of reasoning, but it should be emphasized that there is no general opposition between automatic and systematic cognition: in many cases the same answer will be endorsed by both systems. Quick recall and methodical examination agree about the simple anagram and arithmetical problems; equally, if one automatically responds “1066” when asked the date of the Battle of Hastings, one can also arrive at this same response after some methodical examination of the relevant historical research.

Not every problem is open to handling by either system, however: some problems have features System 1 is unable to process. In particular, System 2 is needed for any problem which requires us to represent a hypothetical possibility and then negate it, an operation requiring the intervention of working memory. In the terminology of Strack and Deutsch, who use the label ‘reflective system’ for System 2, “the cognitive procedure of negating can only be executed in the reflective system” (Strack & Deutsch, 2004, 227). The fact that negation demands System 2 involvement will play a significant role in the next section, which discusses the role of various System 2 triggers in producing the Harman-Vogel pattern of intuitions.

---

<sup>11</sup> For instance, even subjects who get it right report the lingering influence of the automatic impression. Stephen Jay Gould reports that despite his knowledge of the right answer to the Linda problem, “a little homunculus in my head continues to jump up and down, shouting at me – ‘but she can’t just be a bank teller; read the description’” (Gould, 1992, 469).

<sup>12</sup> In one of the experiments Stanovich and West devised to show this, the text from the original Linda problem was used in a between-subjects design with over 400 subjects. Pools of higher- and lower-IQ subjects were asked to rate the likelihood of Linda’s being a bank teller -- (b), just in contrast to decoy statements about other pursuits like elementary school teaching, while other pools of higher- and lower-IQ subjects were asked to rate the likelihood of Linda’s being a feminist bank teller -- (c), in contrast to the same decoys. The average rating for (c) was significantly higher than that for (b) in both above- and below-median IQ groups; indeed, the between-group tendency to rate (c) higher than (b) was even more pronounced between the above-median IQ groups (Stanovich & West, 2008).

### 3. Dual process theory and the Harman-Vogel paradox

Dual-process theory is a general theory of judgment; the Harman-Vogel paradox concerns a particular class of judgments having to do with the attribution of knowledge. Dual-process theory could be well-positioned to explain the ascriber's shifting intuitions if the ascriber naturally switches from System 1 to System 2 cognition as he moves from the lax to the stringent side of the paradox. Although the paradox directly concerns the cognition of the ascriber rather than the subject being judged, it should be noted from the outset that certain features of the subject's cognition will generate natural counterparts for the ascriber, insofar as the ascriber must simulate or represent the subject's cognition in the course of evaluating it. For example, if a subject must use System 2 in order to represent a hypothetical possibility and then negate it, then her ascriber must also switch to System 2 as he retraces the subject's cognitive steps. This section starts by examining system shifts in the cognition of the judging subject, and then ascends to consider the related predicament of the ascriber, first on the stringent side of the pattern, and then on the lax side.

Looking again at the earlier table of examples of the Harman-Vogel pattern, one might start by noting that all of the stringent judgments consist in the negation of a hypothetical possibility, a function agreed by all dual process theorists to involve System 2 cognition. To judge that it is not the case that Barack Obama has been assassinated in the last hour, we must first represent the possibility of the assassination in the affirmative and then negate it (Gilbert, 1991; Hasson & Glucksberg, 2006), a task requiring the serial processing characteristic of System 2. Similar operations are required in every one of the judgments in the stringent column of the table.

A broader study of the literature confirms the hypothesis that System 2 processing of one kind or another is always required for the stringent claims in the Harman-Vogel pattern of intuitions.<sup>13</sup> Most often, sentential negation is involved in the shift up to System 2, but there are cases that involve other System 2 triggers. Surveying nine core books and articles on the problem (DeRose, 1996; Douven, 2007; Dretske, 2005; Harman, 1973, 1986; Harman & Sherman, 2004; Hawthorne, 2004; Hawthorne, 2005; Vogel, 1990), I identified 44 examples of the relevant pattern of intuitions.<sup>14</sup> In 40 out of these 44 examples (91%), the stringent claim in the pattern involves

---

<sup>13</sup> To say that a switch to System 2 processing is involved in generating a stringent knowledge assessment is not to say that such a switch is *sufficient* to generate a stringent knowledge assessment; it will be argued in the next section that some further features of the stringent propositions combine with the shift to System 2 to explain our reluctance to ascribe knowledge. Other types of System 2 judgment can figure unproblematically in positive knowledge ascriptions.

<sup>14</sup> Where a single example was presented multiple times by an author within a book or article, only the first instance was counted and analyzed. However, when an example generated by one author was taken up, perhaps with some variation, by another, it was recounted on its first presentation in the second author's

sentential negation. The negation is typically overt, as when Vogel asks in his 1990 paper, “Do you know that George Bush hasn’t had a fatal heart attack in the last five minutes?” -- (1990, 20). In a small minority (10%) of cases involving negation, the presentation of the example does not itself use sentential negation, but drives us to recognize the need to embrace a negated proposition in order to sustain the lax judgment just made. In Hawthorne’s version of the current President case, for example, after testing our lax intuitions on the identity of the President we are simply invited to consider the possibility that “perhaps he died in the last five minutes” (2004, 4). Systematic cognition would already be needed to represent the hypothetical scenario of recent death, and once this scenario is represented, we can immediately see its inconsistency with the lax judgment just made. In the struggle to render our judgments consistent, we find that what we need to know that it is not the case that the President has died in the last five minutes. This is the same sort of negated proposition as was presented explicitly in Vogel’s version of the case: that Hawthorne leaves the work of constructing the negation to the reader does not diminish the cognitive demands this judgment imposes.

In many of the stringent propositions involving the negation of a hypothetical possibility, System 2 processing is needed not only to process the negation, but also to handle further features of the proposition. For example, unusual or exceptional situations like those involving cleverly disguised mules and papier-mâché fakes are resistant to automatic processing, given that System 1 relies on association to familiar past instances, and System 2 is triggered by novel or abnormal cases. Even in the absence of negation, presentation of the numerical odds of the lottery would be expected to trigger controlled processing (Windschitl & Wells, 1996). Although Harman’s original lottery scenario was typically introduced with sentential negation – the stringent proposition spoke of a ticket *not winning* – negation would not be necessary to engage System 2 once the odds of the lottery have been mentioned. So, in his initial presentation of the lottery case, DeRose works without negation, but in a manner that still engages systematic cognition.<sup>15</sup> The stringent but at least superficially affirmative proposition “that your ticket is a loser” is introduced after we have

---

work, partly because it is difficult to distinguish variations on older examples from new examples, and partly because I think that examples of the pattern that seem to resonate with many authors should be counted more heavily. The fact that an example is repeated is some evidence that it does a particularly good job of generating the desired paradoxical pattern of intuitions.

<sup>15</sup> DeRose’s version is an exception to a general pattern: in every other author’s presentation of Harman’s original lottery scenario, in the first presentation of the scenario the stringent proposition spoke of *not winning* the lottery rather than losing it. For several authors later repetitions of the example sometimes spoke directly of loss, but only the first presentation of any example was counted in the statistics.



been reminded that this ticket is “one of 20 million tickets, only one of which is a winner” (DeRose, 1996, 568).

Among the 44 cases of the Harman-Vogel pattern surveyed, 40 contained stringent propositions involving sentential negation; DeRose’s “loser” version of the lottery case is one of the four surveyed cases that did not. Like the DeRose case, however, all of the affirmative stringent propositions have special features that require systematic processing; the three remaining cases will be discussed individually before turning to the question of how we judge the propositions on the lax side of the pattern.

The first of the three remaining affirmative stringent propositions involves explicit disjunctive reasoning, the contemplation of a series of hypothetical alternatives. Hawthorne invites us to consider a case in which it is supposed that one trusts the *Guardian* and the *Times* equally, and one reads in the *Times* that Manchester United has won a match. Although one might readily take oneself to know that United won, and that the *Times* says so, Hawthorne points out that it is harder to accept that one knows the following: “either *The Times* and *The Guardian* correctly reported a Manchester United victory or else *The Guardian* made a mistake about the Manchester United result” (2004, 71). Contemplation of these distinct hypothetical scenarios demands sequential reasoning, so this proposition must be evaluated in System 2 (Evans, 2007b).

Hawthorne also presents a case which he describes as a variant of one presented by Vogel in his (1993). In this case a person has consumed a pill from a bucket of 10,000 pills, one of which induces inversion of one’s blue-green color discrimination (the others are inert). If this subject forms the belief that he is looking at a blue patch when this is so, Hawthorne suggests, “we may well be inclined to count him as knowing that he is seeing a blue patch.” However, Hawthorne suggests, we will resist the suggestion that “he can deduce and come to know that the pill he took was inert” (2004, 4). One might conceivably read this case as involving some form of negation, triggered by the negative prefix: knowing that the pill is inert is tantamount to knowing that it is not the active one, representing and then negating that special case. But in any event it is a routine matter to recognize blue things as blue; distinguishing inactive from active pills, at a ratio of 9,999:1 seems to require consideration of numerical odds and special cases, and consequently systematic rather than automatic cognition.

The final problematic stringent proposition is from Dretske, who points out that it is much easier to accept that there are cookies in the jar than that there are mind-independent objects there (2005, 14). Some form of negation could conceivably be in play here, again in the negative prefix: perhaps conceiving of objects as ‘mind-independent’ requires contemplation and negation of some

form of mind-dependency, or sequentially imagining the object in one's presence and absence.<sup>16</sup> But whether or not this case demands negation or hypothetical reasoning, the theoretical character of 'mind-independence' may already suffice to trigger System 2: it is thought that explicit representation of this sort of abstract concept is not handled automatically (Satpute & Lieberman, 2006).

The lax propositions present a strikingly different profile. Where sentential negation was present in 91% of the stringent propositions, it is *absent* in all but two of the lax propositions surveyed (95%). (The exceptions, which will receive special discussion in due course, are an example of DeRose's about a person's inability to pay off a loan, and Hawthorne's similar safari affordability case (vi) in the chart above.) In contrast to the stringent propositions, the lax propositions are routine in character: thinking about routine matters such as tomorrow's lunch, a sports headline or picking up one's car, one will ordinarily rely on automatic judgment. Memory automatically supplies an answer to questions about the identity of the current President, the spot where one has parked, and routine events in one's upcoming schedule. This initial intuitive answer may be overridden if stakes become high, or we are made to feel particularly self-conscious about our responses, or given the expectation that we will have to justify them in front of others (Lerner & Tetlock, 1999; Simonson & Nye, 1992). Subsequent systematic analysis of the question may result in our reaffirming our original automatic answer, or in our qualifying it in some way, for example by converting it from a categorical claim to a claim about what is highly probable.

So far, the aim has been to explain why it is initially psychologically easier for a person to judge that, say, Barack Obama is now President than that it is not the case that he has very recently died. Before giving an explanation of the rare lax propositions involving negation, it will be helpful to show why it is generally easy for a person to ascribe knowledge of the lax proposition to others, and relatively hard to ascribe knowledge of the stringent one.

One might expect that assessments of the mental states of others would invariably demand higher-level controlled processing. However, there is evidence that routine assessments of the knowledge or belief of others is not inherently taxing, but rather modular and automatic in character (Leslie, Friedman, & German, 2004; Saxe & Kanwisher, 2003; Saxe, Schulz, & Jiang, 2006).<sup>17</sup> It seems that we do have to exert executive control to perform compounded mental

---

<sup>16</sup> Thanks to Jonathan Weisberg for this observation.

<sup>17</sup> One might wonder whether empirical work on the modularity of belief ascription is investigating the same processes that underpin responses to written stories: perhaps what works in face-to-face assessments differs interestingly from what works when we read the type of fictional narratives that figure in epistemology papers. However, it seems that developmental changes and functional deficits affect non-verbal and story-

ascriptions, or ascriptions involving multiple subjects – to grasp that Sarah believes that Bill knew that Susan wanted to leave, for example (Kinderman, Dunbar, & Bentall, 1998). We may have to engage in some degree of controlled cognition to represent the beliefs of a subject when these beliefs conflict with our own (Apperly, Back, Samson, & France, 2008; Epley, Keysar, Van Boven, & Gilovich, 2004).<sup>18</sup> Neither of these conditions presents an obstacle in the Harman-Vogel pattern, where the relevant ascription is simple and restricted to a single subject, and it is understood that we are evaluating the subject on the assumption that the animal is in fact a zebra, the President is still alive, and so on. When a third party believes a proposition one takes to be true, the task of representing it as a belief held by another seems to add no special burden of controlled cognition beyond whatever was necessary to represent the proposition itself (Saxe, Schulz, & Jiang, 2006).

It is now possible to lay out a relatively simple model of the lax ascriptions in the pattern. In a context where it is assumed that Obama is still alive and Lucy is a typical compatriot of his, when we evaluate the claim that Lucy knows that the current US President is Barack Obama, we would do so heuristically, on the basis of our own prior knowledge. For example, we might automatically check the truth of proposition believed by Lucy against our own background view, and associatively compute the acceptability of this predication of knowledge on the basis of its similarity to some prototypical representation of knowing (cf. Goldman, 1993). Given that memory for commonplace public domain facts is typically taken for granted, and given our own agreement with the proposition about the identity of the President, the heuristic evaluation of Lucy's epistemic position will be positive.

Something changes when we ask ourselves whether Lucy knows that it is not the case that Obama has died in the last hour. In evaluating Lucy's doxastic attitude to this negated proposition, we need to represent her cognition as controlled, whether we do so by simulating her thought about that proposition or by theorizing about how she would think if she were to attempt to judge

---

comprehension belief attribution tasks at the same pace and in the same way, strongly suggesting a common underlying process is involved in these tasks. For a detailed empirical defense of the legitimacy of using written narrative scenarios to probe human belief and knowledge ascription capacities in general, see (Saxe, 2006).

<sup>18</sup> There is some controversy over the degree of automaticity and spontaneity of false belief ascription under various conditions. For example, where Ian Apperly has aimed to show that there are significant cognitive costs involved in the ascription of false belief (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006), Adam Cohen and Tamsin German have argued that what is cognitively costly is the preservation of a representation of false belief rather than its initial generation (Cohen & German, 2009). Cohen and German find evidence that false belief ascription can be automatic and spontaneous. Meanwhile, Markus Aichhorn and colleagues found evidence of fully automatic belief ascription under conditions in which participants expected to evaluate a variety of true and false beliefs (Aichhorn et al., 2009); they conjecture that false belief ascription may be more or less automatic depending on the expectations and mindset of the ascriber.

that it is not the case that Obama has died in the last hour.<sup>19</sup> If we are simulating, the systematic character of our own cognition will naturally be projected onto Lucy (Goldman, 2006); if we are theorizing, then in representing her as attempting to reach the type of judgment ordinarily supported only by controlled reasoning, we would naturally and charitably represent her reasoning as controlled (Gopnik & Wellman, 1994). In either case, we need to generate some representation of Lucy's evidential position in order to see whether controlled thinking could allow her to advance from her evidence to this particular conclusion in an unbiased manner. Because she is thinking in System 2, Lucy will be aware of the process of moving from evidence to judgment, and her resulting confidence level will depend on the character of this process. What follows is one possible model of the features of controlled cognition that would naturally lead to a negative answer to the question of whether she knows the stringent proposition.<sup>20</sup>

As in the automatic case, we can see her as retrieving a representation of Obama as President from memory. In the automatic case, the judgment that Obama is President would be made automatically when this representation comes to mind, on the strength of phenomenal features of the representation (Johnson, Hashtroudi, & Lindsay, 1993) and metacognitive monitoring of memory performance. Given the presumed extent of Lucy's relatively recent evidence, she would experience great fluency of retrieval, which will in ordinary low-stakes contexts suffice to generate high confidence (Kelley, 1993; Schwarz, 1998). In the systematic case, the subject will experience greater self-consciousness about the origin of her judgment; when Lucy is wondering (or, simulating: when we are wondering) whether it is not the case that Obama has very recently died, even a convincingly recent memory of him in that role will not produce the same confidence in his continued existence. When we are thinking systematically, we are aware not only

---

<sup>19</sup> There is a lively current debate over whether our natural 'mindreading' abilities are better understood in terms of simulation or theorizing, although it is fair to say that there is also some recent convergence between the two sides, with elements of simulation now figuring in the "theory-theory", and vice versa. For a presentation of the major positions in the debate, and some evidence of the trend towards common ground, see for example (Goldman, 2006). For the purposes of the present paper I will attempt to stay neutral on these particular issues.

<sup>20</sup> The model presented here relies on Marcia Johnson's Source Monitoring Framework, one mainstream account of judgment; other somewhat similar explanations of the natural shift in intuitions may be constructed in a variety of other psychological models of judgment. It is not within the scope of the present piece to defend Johnson's framework over its rivals. It should also be stressed that in addition to our natural tendency to become more self-conscious about sources of information when we are thinking in a controlled fashion, controlled cognition may be subject to further pressures. For example, one might conduct a controlled assessment of Lucy's predicament by deliberately consulting one's favored epistemological theory. But such an assessment, whether negative or positive, would not offer independent confirmation of that theory of the type typically sought by those whose epistemological theories are sold on their capacity to systematize intuitions.

of the product of our judgment (yes, Obama is President), but also of the process, where the relevant process is interestingly different from the process underpinning automatic judgment. Under systematic cognition, memory-based judgments are recognized as based on past learning, and what is recalled is dated with more precision, with the aid of controlled processes such as deliberate reconstruction of the temporal context of what is recalled (“when did I see him last?”) (Johnson, 1997; Johnson, Raye, Schacter, & Scarry, 2000; Johnson et al., 1993). Once brought into focus in this fashion, the past tense character of what is recalled is evidently problematic as a basis for responding to questions about what has happened very recently. The conspicuously past tense character of controlled memory figures in many instances of the Harman-Vogel pattern of intuitions; in other instances, comparable features of other information sources play a parallel role. For example, in Dretske’s cookies/papier mâché cookie-replicas case, what we automatically judge to be a cookie on the basis of vision cannot be so judged when we shift into System 2 and become concerned with ruling out the possibility of a visually indistinguishable replica. What automatic judgment could accept unconditionally, controlled judgment can accept only subject to qualification, conscious of certain limitations of its evidence, where the evidence available to controlled judgment is not necessarily the same as the evidence available to automatic judgment (Schwarz, 1998, 2004).

This is not to suggest that the less guarded automatic judgment is necessarily inferior to the more reserved controlled judgment. If our System 1 judgments generally hit the mark they may well have a claim to count as manifesting knowledge, even though some of the key epistemic virtues of these judgments – for example, their ecological validity or reliability – are not guaranteed to be first-person accessible to the judging subject, either at the moment of heuristic judgment, or at a later moment when she is reflecting on her past judgments, and wondering about their epistemic merits. The fact that controlled cognition puts a premium on considerations that could be explicitly cited as reasons by the judging subject does not necessarily give its deliverances a better claim to the status of knowledge. Indeed, the fact that automatic and controlled judgment operate rather differently does not entail that there is any difference in the set of features in virtue of which a judgment of either type counts as knowledge; that issue remains an open question here.

Because Lucy in her controlled mode of cognition is presented as making her judgment on the basis of evidence that only renders it very probable that Obama has not just died – she does not have our God’s-eye perspective on the case, in which his continued existence is stipulated, or

understood to be stipulated<sup>21</sup> – she is most charitably represented as making the guarded judgment that it is very probable that he is not dead, rather than the categorical judgment that he is not dead. It is possible for us to represent her making the categorical judgment that Obama is still alive even as she reflects on the dated character of her evidence in the controlled mode of thought. However, in order to do so we would have to see her as engaging in some form of biased cognition, such as wishful thinking (Kunda, 1990), a move that would serve equally well to block a positive evaluation of Lucy’s judgment as manifesting knowledge of this proposition.<sup>22</sup>

The particular type of evidential position that Lucy enters under controlled cognition is important in generating our negative assessment: we have no trouble appreciating that an ordinary subject who engages in some calculation can know that 435 minus 287 is 148, or that it is not 135. We have to think systematically in the course of such an evaluation, say by running through the computation ourselves to check the truth of the proposition believed. But having done so, we do not ordinarily hesitate to ascribe knowledge rather than mere belief here, because we can represent any ordinary adult as having comparable systematic means at her disposal. Equally, when we try to assess whether Lucy knows that it is not the case that Obama is not the President (where it is stipulated that Lucy is adept at logic), System 2 thought is certainly required to comprehend the problem, but if we proceed by imagining Lucy first cancelling the double negation, we can convert this – or, equally well, see her as converting it – into a System 1 problem.<sup>23</sup> We recognize that Lucy could see this as equivalent to asking whether the President is Obama, and then relax into System 1 to answer that easier question automatically, on the basis of prior knowledge.

---

<sup>21</sup>Indeed as readers of the case, once we recognize the stipulation about Obama’s continued existence for what it is – a stipulation applicable to the fictional scenario we are assessing, and not a guarantee of real world outcome – we may ourselves find it difficult simply to accept the categorical proposition about Obama’s not having died, as opposed to the qualified proposition about what is very highly probable. Kyburg’s original version of the lottery paradox concerned rational acceptance rather than knowledge: to handle that version of the paradox it helps to understand why it can be difficult to achieve categorical belief in a very high-probability proposition in a controlled mode of cognition. That there should be some role for a psychological account here is perhaps to be expected given that it has been established that the paradox cannot be solved by appeal to any purely logical or probabilistic characteristics of the propositions involved (Douven & Williamson, 2006). Extending the dual-process strategy to the rational acceptance version of the paradox is however a project for another occasion.

<sup>22</sup> One of the main themes of (Hawthorne, 2004) is that the Harman-Vogel paradox is an embarrassment for reliabilism, because we are reluctant to grant knowledge of stringent propositions even when a judgment of such a proposition is very likely to be right. If one is more likely to be right about lottery ticket loss than about some mundane matter judged on the basis of sense perception, Hawthorne argues, the reliabilist will not be able to explain why judgments of the latter type can come across as manifestations of knowledge when judgments of the former type do not. However, this line of criticism against reliabilism is blocked if an accuracy-compromising process is required to elect the categorical judgment that the ticket has lost over the graded judgment that the ticket was highly likely to have lost.

<sup>23</sup> Thanks to Roy Sorensen here.

We do have a well documented tendency to relax into System 1 whenever possible – a tendency Stanovich dubs “Rule 1 of the Cognitive Miser” – across a very wide range of different contexts (Evans, 2006; Stanovich, 2009; Tversky & Kahneman, 1974).

At least in some cases, it is possible for us to reverse an originally negative epistemic evaluation by thinking harder about further evidence available to the subject, and reasoning through to the conclusion that the subject of the case actually would have adequate evidence for his judgment, even in a controlled mode of cognition. This looks like what happens if we follow Vogel’s reasoning on the zebra case, where the immediately highlighted visual information was insufficient to discriminate a zebra from a cleverly disguised mule, but further information about the background facts about zoos was supposed to reassure us on the question of trickery. The fact that the exonerating evidence takes time to retrieve and process in this case would explain why our initial impression was negative and yet our considered judgment was positive, assuming we are satisfied with Vogel’s reasoning about the background facts. There are of course many other ways of using controlled cognition to override initial epistemic evaluations, including the deliberate application of an epistemological theory.<sup>24</sup>

We are now in a position to discuss the two lax propositions which include a negation, both of which emerge as variations on Harman’s original lottery case.<sup>25</sup> In DeRose’s version of the lottery case, you find it hard to say that you know that your one-in-20-million ticket has lost, and yet easy to say that you know you will not be able to pay back a certain loan by the end of the year (DeRose, 1996). Hawthorne’s version of the case is similar in spirit, but his presentation is more detailed, and will therefore serve as an initial focus for discussion:

Suppose someone of modest means announces that he knows he will not have enough money to go on an African safari this year. We are inclined to treat such a judgment as true, notwithstanding various far fetched possibilities in which that person suddenly acquires a great deal of money. We are at some level aware that people of modest means buy lottery tickets from time to time, and very occasionally win. But despite all this, many normal people of modest means will be willing, under normal circumstances, to judge that they know that they will not have enough money to go on an African safari in the near future. (Hawthorne, 2004, 1-2)

---

<sup>24</sup> An epistemologist well rehearsed in a certain type of anti-skeptical theory may even reach a point at which positive assessment of certain frequently encountered stringent propositions becomes automatic: at some point the words “cleverly disguised mule” may automatically generate a positive rather than a negative response in a suitably habituated epistemologist. Variation of examples is useful in part for this reason.

<sup>25</sup> Further cases are not hard to generate: to take one example suggested by an anonymous referee, it seems easy to credit Lucy with the knowledge that Biden is not the President (as long as worries about recent assassination have not been raised). That example would be handled in the same manner as the Hawthorne safari case discussed below.

One might attempt to come up with a model of judgment in which no controlled processing is required to accept that this nameless “someone of modest means” will not have enough money to go on an African safari this year. As a starting point, one could note that the concept “modest means” includes an upper bound: part of what it is to be a man of modest means is to be unable to afford such luxuries as extravagant trips. If the ascription is made simply by activating the ‘modest means’ stereotype or concept (which includes positive and negative elements) and assessing its fit with the negative predicated category (‘unable to afford an African safari in the near future’), then one might think it is possible to execute this ascription fully automatically. For example, Deutsch, Gawronski and Strack have argued that although propositional negation ordinarily requires working memory, in cases where negation is already included in the associative baggage of a term, “evaluating negated expressions can be driven by memory retrieval instead of the application of the procedure to negate” (Deutsch, Gawronski, & Strack, 2006, 393).

A more complex, although probably empirically better grounded<sup>26</sup> explanation of the laxity of the safari ascription would characterize our thinking about such matters as ordinarily involving a mixture of System 1 and System 2 cognition. For example, we could heuristically generate a representation of what a stereotypical person of modest means can ordinarily afford, and then judge that the safari vacation will exceed that range, where only this second operation is controlled or analytic. Because different versions of dual-process theory have somewhat different accounts of the integration of the two systems, it would be very difficult to supply a generic dual process story at this point. A number of different accounts might be consulted, but for our present purposes the “satisficing” model developed by Jonathan Evans is particularly easy and effective.

According to Evans, the heuristic system by default generates a single model of the ‘most probable or believable state of affairs’, whose analytic evaluation is constrained by a satisficing principle, ‘a fundamental bias in the analytic system to work with the representation it has unless there is good reason to give it up’ (Evans, 2006, 379). One possible motivation to give up the heuristically generated representation of what someone with modest means can afford would be the recognition that our answer to the safari affordability question is inconsistent with our answer

---

<sup>26</sup> Although it would yield a way of characterizing this lax judgment as purely heuristic, it may be a stretch to apply the Deutsch theory of automated negation in this case. Deutsch’s work focused on formulaic expressions like “no way”, and the problematic proposition in the Hawthorne example involves clear sentential negation. It remains an interesting empirical question how far Deutsch’s theory can be extended, and whether it might cover cases such as this one.



to the fair lottery loss question. Until we hit such a problem, however, it will be natural for us to make judgments on the basis of our simplified mental model.<sup>27</sup>

Even if we are at some level aware that a normal person's ability to afford something in the next year is a function of many factors, the availability of a stereotype of what the 'person of modest means' can most likely afford lets us reach an intuitively compelling answer to the question of what such a person can afford without going through the burdensome exercise of figuring out what all those factors are and compounding their probabilities and payoffs. Given appropriate information and adequate incentives we might construct a more fine-grained model or set of models of what the subject could afford under various circumstances; given different motivations – for example, pressure towards immediate action<sup>28</sup> – we would be much less likely to overturn our initial heuristic impression (Kruglanski & Webster, 1996).

The satisficing model of cognition will not enable comparable shortcuts on the problem of loss in a fair lottery, unless that prospect is represented in a way which opens the door to heuristic cognition. On that frontier, Hawthorne notes the following:

It is relatively easy to get ourselves in the frame of mind where we reckon ourselves to know that we will not win the New York State lottery each of the next thirty years (even if we expect to buy a ticket each year). Just ask people. They will happily claim to know that *that* will not happen. (2004, 20)

Meanwhile, Hawthorne observes, if by 'using normal statistical methods' we compute the odds of winning thirty such lotteries to be  $1/n$ , 'We might point out to someone that if he had a ticket in one great big lottery with  $n$  tickets, he would not reckon himself able to know that he would lose in *that* case. Intuitions would then switch.' (ibid.) The switch in intuitions here suggests that our ordinary assessments of compound outcomes do not always reflect 'normal statistical methods': in contemplating the dynamic scenario of consistently winning one random lottery after another, we circumvent the mathematical work of representing the compounded probabilities, and judge the problem in some other fashion. One relevant heuristic process would involve probabilistically

---

<sup>27</sup> The Evans analysis applies equally to DeRose's negated lax proposition: if we have a heuristically generated representation of our anticipated spending power for the year, we can take it as an unquestioned input to our more systematic calculations concerning our capacity to repay the loan, at least until motivated to reconsider.

<sup>28</sup> Controlled processing takes time; when one faces immediate pressure to act there is a greater tendency to judge on the basis of heuristic impressions. Arguably, the connection between action and heuristic thought explains some of the correlations between practical circumstances and knowledge ascription noted by Hawthorne in his (2004). For detailed criticism of such claims on the relation between knowledge and action see (Douven, 2008).

inaccurate but psychologically easy heuristic models of randomness, which exclude obvious regularities, making it seem, for example, that the coin toss sequence HHHHTTTT is much less probable than THHHTHTT (Kahneman & Tversky, 1972). Thanks to these models, the judgment that one simply cannot win 30 fair lotteries in a row has a certain natural strength; however, with sufficient cognitive effort it is possible to reason through to the contrary systematic judgment that it is just very highly unlikely that one will enjoy 30 consecutive lottery wins.

#### **4: The implications for closure**

One benefit of the dual-process analysis is that it can explain why the Harman-Vogel pattern of intuitions does not really pose a threat to the principle of closure. Assuming that the automatic judgment of a lax proposition is sufficient for knowledge (the possibility that it is not will be discussed in the next section), one might wonder why a given subject cannot simply deduce the stringent proposition from the lax one on the strength of the recognized relation of entailment. What happens is that to attempt to execute the deduction – a paradigmatic controlled process – we must switch into System 2, where we discover obstacles to embracing the stringent judgment. Then, insofar as we continue to recognize the relation of entailment between the lax and the stringent proposition, and continue in a controlled mode of cognition, we will also be unable to make that particular lax judgment. While I am dwelling on my inability to rule out the possibility that my car has been stolen, and engaging in controlled cognition on the question of its whereabouts, I can no longer automatically judge it to be parked on Avenue A in the manner that might earlier have manifested knowledge. Equally, if thinking about lottery wins has pushed me to construct a more detailed set of models to cover what a person of modest means can afford, then I can no longer make my simple single-model judgment that this person will be unable to afford a safari trip.

Having shifted into the controlled mode of cognition, it is true that I may be able to recall the fact that I initially judged that my car was on Avenue A before I started to worry about the possibility of auto theft. It would be unproblematic for me to use this prior judgment as the starting point of my deduction if I could know now that my earlier judgment constituted knowledge; however, knowing that P does not in general guarantee that we are positioned to know that we know that P (Williamson, 2000). In particular, the reasons why the automatic judgments in the pattern manifest knowledge would not be predicted to be accessible to the subject who is engaged

in controlled cognition about the propositions on the stringent side of the pattern, for reasons discussed in the last section. Meanwhile, the shift from judging that  $p$  to finding oneself unable to judge that  $p$  could naturally be felt as a loss of knowledge, perhaps in part because we can be unaware that we have shifted from one cognitive mode to another (Cary & Reder, 2002). If we are accustomed to switching from accepting to withholding on a proposition only when we encounter new information about the world, the fact that we find ourselves making such a switch to withholding upon encountering the stringent proposition could readily give us the sense that we have come to know less.<sup>29</sup>

Losing one's capacity to judge that  $p$  does not necessarily entail losing one's knowledge that  $p$ : it might still be true of me that I know that my car is on Avenue A even if I am presently unable to convert this knowledge into explicit judgment.<sup>30</sup> But it is not enough to have knowledge of a proposition in order to execute a competent deduction in which it figures as a premise: because a deduction involves a series of explicit judgments, one needs to be able to judge that  $P$  in order to use  $P$  as a premise in a deduction.

We are familiar with the idea that the range of propositions one can rationally embrace can shift as one comes to learn new facts about the world. What the dual-process account adds is that the range of propositions one can rationally embrace in explicit judgment can also shift as we move from one cognitive mode to another, from System 1 to System 2. Some propositions – for example, that  $9 \times 11 = 99$  – can be made in either cognitive mode; others can be safely made only in one mode or the other. The question of whether Lucy's judgment that Obama is President manifests knowledge will depend in part on whether she is presently thinking about that proposition in an automatic or controlled manner. That she could at some point make the automatic judgment safely and without bias does not entail that she could make the controlled judgment unproblematically:

---

<sup>29</sup> Thanks to Jonathan Weisberg for discussion of this point. Note that this is a restricted error theory, applying only to the sense of knowledge loss that may appear when we attempt to deduce the stringent proposition from the lax one. On the dual process account, there is no mistake in the initial core intuitions of the paradox – that one knows the lax proposition and does not know the stringent one.

<sup>30</sup> For this observation, and for the general direction of this section of the paper, I am indebted to Tim Williamson. Alternatively, one might attempt to argue that the shift into System 2 *does* destroy one's knowledge of the lax proposition, if only temporarily. On such a view closure is still preserved, although it may become an awkward matter to say whether someone knows a proposition that she is not occurrently judging in either cognitive mode. Taking such a view might oblige one to embrace a theory of knowledge or belief such as Norbert Schwarz's construal theory, on which occurrent judgment or manifest knowledge is psychologically real and dispositional belief or knowledge is a derivative construct (Schwarz, 2007). One can remain neutral on the truth of construal theory by emphasizing instead that what is lost is one's capacity to make the explicit judgment that would be needed for the deduction in the antecedent of the closure principle.

once she has responded to worries about recent assassination by focusing on the temporal features of her memorial evidence, she can only elect the categorical judgment over the graded one by force of wishful thinking or some comparably biased process. Equally, as long as I am thinking about lottery wins or traffic accidents, certain explicit judgments will be out of reach for me: given my evidence in the controlled mode of cognition I should at most judge it to be very probable that I am meeting my friend for lunch tomorrow or teaching epistemology in the fall.

This notion that a single proposition might manifest knowledge if judged in one manner but not in another dates back at least to Firth's distinction between propositional and doxastic justification (Firth, 1978). Epistemologists unaccustomed to thinking in terms of dual-process theory may want to consider cases involving other sorts of changes in cognition. Consider a jury member who is inappropriately eager to vote for conviction, in a case in which the prosecution has presented excellent physical evidence and rather weak testimonial evidence of the defendant's guilt. If this juror chooses to reflect on the physical evidence – the DNA results, the fingerprints and so on – his judgment that the defendant is guilty could count as manifesting knowledge. His judgment of the same proposition might fail to count as knowledge if he chooses to make it while thinking only about the testimonial evidence presented, brushing aside the gaps and faults in what he heard. To assess whether his guilty verdict manifests knowledge, one needs to consider not only the information available to the subject, but also about the manner in which this information was handled, the actual basing of the subject's belief.

The distinction between knowledge and explicit judgment enables us to see that there is something missing in Hawthorne's formulation of the rule of single premise closure:

If one knows P and competently deduces Q from P, thereby coming to believe Q, while retaining one's knowledge that P, one comes to know that Q. (2005, 29)

The competent deduction clause states that one is to deduce Q from P; it does not say whether one's starting point must be one's knowledge that P, or could be a mere judgment that P. Reading the principle the first way, the peculiarity of the Harman-Vogel pattern is that the relevant knowledge will be inaccessible to explicit judgment in the controlled mode of cognition, so even if the knowledge that P is not lost, no competent deduction can start from a judgment that P which itself manifests this knowledge. On this reading, it will not be psychologically possible for us to realize the antecedent of the closure principle for judgments in the Harman-Vogel pattern, so the pattern presents no threat to closure. Reading the principle the second and weaker way, the deduction

must begin from a mere judgment that P, which itself need not count as knowledge. For example, we imagine Lucy initially knowing that Obama is President, and being able to judge automatically that he is. Then in the course of wondering whether he has not been assassinated in the last hour, she attempts to construct a deduction of this proposition from an explicit judgment that he is now President, where this explicit judgment – if what has been said in section 3 is correct – will not itself manifest knowledge.<sup>31</sup> Perhaps the construction of such a less-than-knowledgeable judgment by itself suffices to quash Lucy’s claim to have retained knowledge of the original premise; if so, the antecedent of the closure principle is not realized. But if it does not – if it is possible to make a poorly founded judgment of a proposition one knows without losing knowledge of that proposition<sup>32</sup> – then Lucy can satisfy the antecedent of Hawthorne’s closure principle by competently deducing the conclusion that the President has not been assassinated from the premise that he is alive. However, we have clear reason to doubt that her conclusion deserves the status of knowledge here, given that she was able to reach it only by violating epistemic norms to get the needed premise. This weakened reading of the closure principle is unappealing; the principle should be amended to rule out deductions starting from judgments that fail to manifest knowledge of the judged propositions, if such judgments are in fact possible while knowledge is retained.

## 5. Epistemological consequences

So far, this paper has had the primary aim of explaining why certain judgments feel natural to us. It is another matter which of these judgments are right. It seems there are various possible epistemological interpretations of dual-process treatment of the Harman-Vogel pattern of intuitions. There are aspects of the dual-process account that may appeal to skeptics: when stringent second thoughts tend to override lax first impressions in assessments of probability, we do not hesitate to endorse the more stringent judgments. If something comparable is happening in epistemic assessment, then this could be seen as further evidence of the inadequacy of our lax

---

<sup>31</sup> What grounds this controlled judgment cannot be her newly self-conscious reliance on memory; it could be something like: “Obama simply must be alive; the country needs him now.”

<sup>32</sup> The phenomenon of judgment that fails to manifest knowledge is the counterpart for theoretical reason of the phenomenon of *akrasia* or weakness of the will for practical reason. On the possibility of theoretical *akrasia*, see (Tenenbaum, 1999). It is noteworthy that one of Tenenbaum’s examples concerns a clash between the heuristic reasoning that makes the gambler’s fallacy appear plausible and the sometimes psychologically less compelling controlled reasoning that proves otherwise.

positive impressions. On the other hand, contextualists might hope that dual process theory could support the contention that there are different standards for the assessment of knowledge, use of which varies systematically across different circumstances. In what follows I discuss possible skeptical and contextualist interpretations of the dual-process account, note some reservations one might have about each, and discuss some alternatives.

Skepticism is sometimes sold as an explanation of the Harman-Vogel pattern of intuitions. One reason why this pattern of intuitions has been taken by skeptics to support their stance is that, as Vogel pointed out, for almost any empirical proposition one might claim to know, some stringent proposition can be generated to undercut one's capacity to make such a claim.<sup>33</sup> Having noted this, the skeptic could insist that clashes between the more positive judgments produced by System 1 and the more cautious judgments of System 2 are appropriately settled in favor of the latter. If Lucy's categorical judgment that Obama is President depends on her not having scrutinized her memorial evidence, then so much the worse for the claim that she has knowledge of any such categorical proposition. Once we recognize that she would not rationally make the categorical claim if she were thinking about her evidence in a controlled manner while contemplating the real possibility of Obama's recent demise, the skeptic will urge us to fault her for making the categorical claim in any circumstances: surely the judgments of those who are thinking about their evidence in a controlled manner are to be preferred to the judgments of those who are relatively oblivious. Furthermore, given that automatic cognition leads us into such traps as violations of the axioms of probability, what reason do we have to trust its deliverances about knowledge?

In response to the skeptic, one might challenge the suggestion that controlled cognition is invariably preferable to automatic cognition. On the Linda problem there is no question that those who complied with the conjunction rule generated a better ranking of likelihood, but in some

---

<sup>33</sup> It might initially seem that the sort of undercutting posited by the dual-process account is not strong enough for the skeptic: the skeptic wants us to suspend belief on the question of whether Obama is President, rather than taking the milder step of judging that it is only highly likely that Obama is President. But the same basic strategy employed to shift from the categorical judgment to the graded one can be applied again to diminish confidence in the graded proposition. The skeptic who disputes the credentials of System 1 cognition will be pleased to note that System 2 typically takes the products of System 1 as input. If one's controlled assessment that it is highly likely that Obama is President itself depends on a range of automatic judgments (about the reliability of TV news reports, and so on), then a general suspension of trust in automatic judgment will lead in short order to very widespread suspension of belief. Or, more directly, instead of raising a series of mundane worries about mortality, and then about the reliability of TV news, the skeptic can raise more radical worries about brains in vats.

situations it seems System 1 can outperform System 2 (e.g. Albrechtsen, Meissner, & Susa, 2009; Gigerenzer, 2008; Wilson & Schooler, 1991). Even in domains such as probability judgment, where controlled cognition is generally an improvement on automatic cognition, we naturally continue to suffer certain deficits even after switching to the systematic mode of cognition. Indeed, with some problems – for example, noncausal base rate problems – performance actually worsens in individuals thinking more systematically (Stanovich & West, 1998). For such problems, ordinary individuals need formal training in the adoption of the correct systematic rules; without such training, more systematic cognition only leads us further astray. Meanwhile, reflecting on the character of one’s thinking on a given problem exposes one to the epistemic hazards of introspection without necessarily improving one’s insight into the original problem (Kornblith, 2005; Kornblith, 2010). Establishing that a certain type of judgment is produced by systematic cognition does not on its own settle the question of whether it is the right answer to the problem, or even whether it is an improvement over the corresponding heuristic judgment.

Meanwhile, the skeptic should also worry that even if the default rules naturally employed in our systematic epistemic assessments often deliver skeptical results, it’s not obvious the *right* rules will do so (after all, the default rules employed in systematic assessments of probability endorse ignorance of noncausal base rates). If the kind of thinking we engage in when contemplating complex and atypical scenarios really is controlled cognition, then it should be sensitive to the deliberate adoption of epistemological theories. Just as systematic statistical judgments are sensitive to training, epistemological judgment may become more resistant to skeptical pressure for those practiced in applying an explicit non-skeptical epistemological theory.<sup>34</sup> The skeptic cannot assume that deliberately adopted theories should ultimately have to be skeptical in character; indeed, it is not obvious that meaningful rules for epistemic assessments can be constructed at all without some guidance from our initial heuristic assessments, assessments which don’t look skeptical at all. In summary, if the dual process account of the Harman-Vogel pattern of intuitions is correct, then the skeptic cannot claim support from this pattern of intuitions without an independent argument to establish that skeptical controlled judgments are epistemically superior to non-skeptical controlled judgments and to automatic judgments. Unless one is already convinced of the merits of skepticism, there is no reason to accept a skeptical story about this pattern of intuitions, and to the extent that skepticism has been

---

<sup>34</sup> On the malleability of epistemic rules and the corresponding intuitions they produce, see (Gendler & Hawthorne, 2005).

motivated by its promise to explain this pattern of intuitions, the motivation for skepticism is undercut by the availability of the dual-process account.

One might expect the contextualist to be in a better position here: where the skeptic downplays the positive assessments made in System 1, the contextualist need not favor one system at the expense of the other. The distinctive claim of contextualism is that the very same subject can rightly be seen as knowing that  $p$  by one ascriber and as failing to know that  $p$  by another. If we recognize both automatic and controlled epistemic ascriptions as legitimate, we have a way of understanding this variation: perhaps the automatic and controlled modes of epistemic assessment would naturally pick out different relations. Of course contextualist theories commonly have more than two settings for knowledge ascriptions, but a more fine-grained account could rely on a more subtle dual-process theory like Evans's satisficing theory, in which judgments are made on the basis of heuristically generated models which are themselves judged by analytic criteria of varying stringency. On an account of this sort, the perceived adequacy of a mental model would naturally depend on such familiar contextualist conditions such as the range of alternatives raised to saliency in conversation.

Some caution is needed in setting up the sort of case that would best favor a contextualist appropriation of the dual-process account. In particular, it would not help to consider a case in which a subject who is making a System 2 judgment is being evaluated differently by two ascribers, one of whom is making a positive System 1 evaluation and the other a negative System 2 evaluation. Because the subject is thinking in a controlled fashion, controlled cognition is required even to represent his thinking accurately: discrepancy between the ascribers in such a case would be most naturally understood as the product of a failure on the part of the lax ascriber to appreciate the subject's real situation. The contextualist would be better off focusing on cases in which a subject who is making a System 1 judgment – say, Smith automatically (and correctly) judging that his car is on Avenue A, where he remembers parking it an hour ago – is evaluated as knowing by an ascriber who is judging him heuristically and evaluated as failing to know by an ascriber who is judging him in a controlled fashion. The contextualist will contend that increases in stringency are natural for ascribers who begin to think in a controlled fashion about the cognition of others. Once I, as an ascriber, start to worry about car theft, it is natural for me to deny that Smith's automatic judgment about the location of his car really constitutes knowledge.

Here an invariantist might object that even if it is natural for worried ascribers to become more stringent, this is not necessarily a good thing. We may grant the legitimacy of automatic and



controlled epistemic assessments without accepting the correctness of the particular controlled assessments that seem to support contextualism; perhaps in these cases our natural tendency towards increased stringency is inappropriate,<sup>35</sup> and a better exercise of our controlled cognition would result in our endorsing Smith's automatic judgment as fully acceptable. In any event, the dual-process explanation of the Harman-Vogel pattern of intuitions increases the burden of argument on the contextualist. If contextualism is marketed chiefly as a way of explaining this pattern of intuitions, the availability of an invariantist version of the dual-process account would pose a genuine threat to contextualism.

One might think that interest-relative or subject-sensitive invariantism (henceforth SSI) would be better positioned to make use of the dual-process account, because of its focus on the judging subject. The advocate of SSI contends that 'knows' always denotes the same relation, but that it is a relation in which the interests and concerns of the subject play a role unrecognized in traditional epistemology. According to the dual process account of reasoning, one's current cognitive mode can make a difference to which propositions one may rationally embrace. For example, it might be acceptable for Smith to (1) judge automatically that his car is on Avenue A, but unacceptable for him to (2) make a controlled judgment that his car is on Avenue A when he is thinking about the possibility of theft. The challenge for the advocate of SSI would be to establish that the difference between (1) and (2) here is not itself best captured in terms favorable to the traditional intellectualist invariantist, who argues that only the subject's state of belief and the truth-relevant features of his judgment make a difference to whether any subject knows or fails to know a given proposition.<sup>36</sup> If controlled cognition in circumstances like Smith's would lead a rational thinker to the graded judgment that it is very likely that his car is on Avenue A, then Smith's categorical judgment (2) becomes possible only if he is thinking in a less-than-rational manner, perhaps suffering from haste, distraction or wishful thinking ("my car simply must be

---

<sup>35</sup> For example, Hawthorne and Williamson have argued that we have a natural psychological tendency to overestimate the likelihood of possibilities of error when these are made salient (Hawthorne, 2004; Williamson, 2005). It has also been argued that our natural tendency towards increased stringency is better explained by the bias of epistemic egocentrism, a tendency to evaluate others as though they shared our privileged concerns (Nagel, forthcoming). Epistemic egocentrism is thought to be the main bias of human mental state ascription (Birch, 2004; Birch & Bloom, 2007; Epley et al., 2004; Keysar, Lin, & Barr, 2003; Nickerson, 1999). When we are thinking in a controlled fashion, this bias would guide us to evaluate others as though they were also engaging in controlled cognition, even when we are explicitly aware that they are judging automatically.

<sup>36</sup> The expression "truth-relevant" is Keith DeRose's, and is meant to encompass all factors that "affect how likely it is that the belief is true, either from the point of view of the subject or from a more objective vantage point." (DeRose, 2009, 24)

there – I need to get home quickly”). But because such conditions lower the accuracy of our judgment, the traditional intellectualist invariantist has no difficulty explaining how (1) could manifest knowledge while (2) fails to do so. On the dual-process account, there is no need to posit stakes as an independent factor in knowledge: because changes in stakes affect the accuracy of cognition, our theory of knowledge can explain shifts that correlate with stake changes simply in terms of the traditional intellectual factor of accuracy.

If the dual process account of the Harman Vogel paradox is correct, the explanation of why we make different judgments as our stakes or concerns change does not lie in any surprising linguistic or metaphysical features of “knows”, but in some basic features of human judgment. It remains an open question exactly how the two types of judgment identified by dual process theory deserve to count as manifesting knowledge, and more specifically, whether it is the same set of virtues that make a judgment of either the automatic or the controlled variety count as a manifestation of knowledge. A full defense of traditional intellectualist invariantism would require much more detailed treatment of issues such as these. But even if it remains unclear just how far we should trust our automatic epistemic assessments, or how we should control our controlled epistemic assessments, we have gained some ground in recognizing the existence of a distinction between the ways in which these assessments are naturally conducted: we are now better positioned to understand the bearing of natural epistemic intuitions on epistemological theory. Whether or not self-consciousness generally improves our judgment, self-consciousness about the origins of our epistemic intuitions can doubtless improve our epistemology.<sup>37</sup>

---

<sup>37</sup> For discussion and comments, I am grateful to Keith Frankish, Mikkel Gerken, Diana Raffman, Andrew D. Ross, Roy Sorensen, Jason Stanley, Keith Stanovich, Sergio Tenenbaum, Jonathan Weisberg, Tim Williamson and two anonymous referees. Earlier versions of portions of this material were presented to audiences at the April 2009 Methodology Conference at the Arché Centre of the University of St. Andrews, the 2009 Canadian Philosophical Association Congress in Ottawa, and the 2009 Thought Experiments Workshop in Toronto. Thanks to participants at all three places for helpful feedback. Thanks also to the Social Sciences and Humanities Research Council of Canada for funding this research.

## References

- Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention. *Journal of Cognitive Neuroscience, 21*(6), 1179-1192.
- Albrechtsen, J., Meissner, C., & Susa, K. (2009). Can intuition improve deception detection performance? *Journal of Experimental Social Psychology, 45*(4), 1052-1055.
- Alter, A., Oppenheimer, D., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology-General, 136*(4), 569-576.
- Apperly, I., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition, 106*(3), 1093-1108.
- Apperly, I., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17*(10), 841.
- Birch. (2004). Understanding children's and adults' limitations in mental state reasoning, *Trends in cognitive sciences* (Vol. 8, pp. 255).
- Birch, S., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*(5), 382.
- Booth, J., Hall, W., Robison, G., & Kim, S. (1997). Acquisition of the mental state verb "know" by 2-to 5-year-old children. *Journal of Psycholinguistic Research, 26*(6), 581-603.
- Cary, M., & Reder, L. (2002). Metacognition in strategy selection: Giving consciousness too much credit. In P. Chambres & M. Izaute & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 63-78). New York: Kluwer.
- Christensen, D. (2004). *Putting logic in its place: formal constraints on rational belief*. New York: Oxford University Press.
- Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness* (pp. 1-40). Hove, UK: Psychology Press.
- Cohen, A., & German, T. (2009). Encoding of others' beliefs without overt instruction. *Cognition, 111*(3), 356-363.
- Cohen, S. (1988). How to be a Fallibilist. *Philosophical Perspectives, 91*-123.
- Cohen, S. (1999). Contextualism, Skepticism, and the Structure of Reasons (Volume 13: Epistemology). *Nous-Supplement: Philosophical Perspectives, 13*, 57-89.
- Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning, 14*(2), 182-199.
- De Neys, W. (2006). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59*(6), 1070-1100.
- DeRose, K. (1995). Solving the Skeptical Problem. *Philosophical Review, 104*(1), 1-52.
- DeRose, K. (1996). Knowledge, Assertion and Lotteries. *Australasian Journal of Philosophy, 74*(4), 568-579.
- DeRose, K. (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context, Volume 1*. New York: Oxford University Press.
- Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology, 91*(3), 385.
- Douven, I. (2007). A Pragmatic Dissolution of Harman's Paradox. *Philosophy and Phenomenological Research, 74*(2), 326-345.
- Douven, I. (2008). Knowledge and Practical Reasoning. *Dialectica, 62*(1), 101-118.
- Douven, I., & Williamson, T. (2006). Generalizing the lottery paradox. *The British Journal for the Philosophy of Science, 57*(4), 755.
- Dretske, F. (2005). The case against closure. In M. Steup & E. Sosa (Eds.), *Contemporary Debates in Epistemology*. Malden, MA: Blackwell.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327-339.
- Evans, J. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.
- Evans, J. (2007a). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

- Evans, J. (2007b). *Hypothetical thinking: dual processes in reasoning and judgement*. New York: Psychology Press.
- Fantl, J., & McGrath, M. (2007). On Pragmatic Encroachment in Epistemology. *Philosophy and Phenomenological Research*, 75(3), 558.
- Firth, R. (1978). Are Epistemic concepts reducible to ethical concepts? In A. Goldman & J. Kim (Eds.), *Values and Morals* (pp. 215–229). Dordrecht: D. Reidel.
- Frank, R., & Hall, W. (1991). Polysemy and the acquisition of the cognitive internal state lexicon. *Journal of Psycholinguistic Research*, 20(4), 283-304.
- Frankish, K. (2009). Systems and levels: dual-system theories and the personal-subpersonal distinction. In J. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 89-107). New York: Oxford.
- Gendler, T., & Hawthorne, J. (2005). The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies. *Philosophical Studies*, 124(3), 331-352.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3(1), 20.
- Gilbert. (1991). How mental systems believe. *The American psychologist*, 46(2), 107-119.
- Goldman, A. (1993). Epistemic folkways and scientific epistemology. *Philosophical Issues*, 271-285.
- Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.
- Gopnik, A., & Wellman, H. (1994). The theory theory. In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.
- Gould, S. (1992). *Bully for brontosaurus: Reflections in natural history*: WW Norton & Company.
- Harman, G. (1973). *Thought*: Princeton University Press Princeton, NJ.
- Harman, G. (1986). *Change in view: Principles of reasoning*: MIT Press.
- Harman, G., & Sherman, B. (2004). Knowledge, assumptions, lotteries. *Philosophical Issues*, 14(1), 492-500.
- Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics*, 38(7), 1015-1032.
- Hawthorne, J. (2004). *Knowledge and Lotteries*. New York: Oxford University Press.
- Hawthorne, J. (2005). The case for closure. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 26-43). Malden, MA: Blackwell.
- Hume, D. (1739/1978). *A treatise of human nature*. New York: Oxford University Press.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin*, 114(1), 3-28.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430-454.
- Kelley, C. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions, *Journal of Memory and Language* (Vol. 32, pp. 1).
- Keysar, B., Lin, S., & Barr, D. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25-41.
- Kinderman, P., Dunbar, R., & Bentall, R. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89(2), 191-204.
- Kornblith, H. (2005). *Knowledge and its Place in Nature*: Oxford University Press, USA.
- Kornblith, H. (2010). What reflective endorsement cannot do. *Philosophy and Phenomenological Research*, 80(1), 1-19.
- Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "Freezing". *Psychological review*, 103(2), 263-283.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
- Kyburg, H. (1961). *Probability and the logic of rational belief*. Middletown, CT: Wesleyan University Press.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. London: Longman.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the Effects of Accountability. *Psychological Bulletin*, 125(2), 255-275.
- Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy*, 74(4), 549-567.
- MacFarlane, J. (2005). The assessment sensitivity of knowledge attributions. *Oxford Studies in Epistemology*, 197-233.

- Nagel, J. (2008). Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy*, 86(2), 279-294.
- Nagel, J. (forthcoming). Knowledge ascriptions and the psychological consequences of thinking about error. *Philosophical Quarterly*.
- Nelkin, D. (2000). The lottery paradox, knowledge, and rationality. *The Philosophical Review*, 109(3), 373-409.
- Nickerson, R. S. (1999). How we know--and sometimes misjudge--what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737-759.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York: Oxford University Press.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988.
- Richard, M. (2004). Contextualism and relativism. *Philosophical Studies*, 119, 215-242.
- Roush, S. (2005). *Tracking truth*. New York: Oxford University Press.
- Satpute, A., & Lieberman, M. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, 1079(1), 86-97.
- Saxe, R. (2006). Why and how to study Theory of Mind with fMRI. *Brain Research*, 1079(1), 57-65.
- Saxe, R., Schulz, L., & Jiang, Y. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, 1(3), 284-298.
- Schiffman, R., & Schneider, W. (1977). Controlled and automatic human information processing: I Detection, search and attention. *Psychological Review*, 84, 1-66.
- Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review*, 2(2), 87.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14(4), 332-348.
- Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, 25(5), 638-656.
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational behavior and human decision processes*, 51(3), 416-446.
- Slovic, S. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological bulletin*, 119, 3-22.
- Stanley, J. (2005). *Knowledge and Practical Interests*. New York: Oxford University Press.
- Stanovich, K. (1999). *Who is rational?: Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum.
- Stanovich, K. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In K. Frankish & J. S. B. T. Evans (Eds.), *In Two Minds: Dual Process and Beyond* (pp. 55-88). New York: Oxford University Press.
- Stanovich, K., & West, R. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, 127, 161-188.
- Stanovich, K., & West, R. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 94(4), 672.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220-247.
- Tenenbaum, S. (1999). The judgment of a weak will. *Philosophy and Phenomenological Research*, 59(4), 875-911.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.
- Tversky, A., & Kahneman, D. (1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological review*, 90(4), 293-315.
- Vogel, J. (1990). Are there counterexamples to the closure principle? In M. Roth & G. Ross (Eds.), *Doubting: Contemporary Perspectives on Skepticism* (pp. 13-25). Dordrecht: Kluwer.
- Vogel, J. (1993). Dismissing skeptical possibilities. *Philosophical Studies*, 70(3), 235-250.
- Williamson, T. (2000). *Knowledge and its Limits*. New York: Oxford University Press.
- Williamson, T. (2005). Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowledge. *The Philosophical Quarterly*, 55, 213-235.
- Wilson, T., & Schooler, J. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2), 181-192.

Windschitl, P., & Wells, G. (1996). Measuring psychological uncertainty: verbal versus numeric methods. *Journal of Experimental Psychology Applied*, 2, 343-364.