Erkenn (2011) 74:399–424
DOI 10.1007/s10670-010-9232-4

ORIGINAL RESEARCH

The Epistemology of Forgetting

Kourken Michaelian

Received: 9 December 2009 / Accepted: 6 June 2010 / Published online: 22 July 2010

Abstract The default view in the epistemology of forgetting is that human memory would be epistemically better if we were not so susceptible to forgetting—that forgetting is in general a cognitive vice. In this paper, I argue for the opposed view: normal human forgetting—the pattern of forgetting characteristic of cognitively normal adult human beings—approximates a virtue located at the mean between the opposed cognitive vices of forgetting too much and remembering too much. I argue, first, that, for any finite cognizer, a certain pattern of forgetting is necessary if her memory is to perform its function well. I argue, second, that, by eliminating "clutter" from her memory store, this pattern of forgetting improves the overall shape of the subject's total doxastic state. I conclude by reviewing work in psychology which suggests that normal human forgetting approximates this virtuous pattern of forgetting.

1 Virtuous Forgetting

Epistemologists have so far paid scant attention to forgetting. This neglect of such a prominent feature of the human memory system (indeed: of all biological memory systems) is natural given (what I take to be) the default view on the epistemic status of forgetting, viz., that forgetting is in general[1] straightforwardly epistemically counternormative, that our propensity to forget is simply an epistemically

[1] The default view should perhaps be understood as allowing that it is epistemically normative for a subject to forget information that she no longer accepts (information that has been left over after belief revision). See Sects. 2 and 3 for discussion of the role of forgetting in eliminating such "left over" information.

K. Michaelian (✉)
Institut Jean-Nicod (CNRS-EHESS-ENS), UMR 8129, Pavillon Jardin,
Ecole Normale Supérieure, 29, rue d'Ulm, 75005 Paris, France
e-mail: kourken.michaelian@ens.fr

unfortunate consequence of the suboptimal design of the human memory system. The thought is conveniently expressed in the language of virtue epistemology: forgetting is in general a cognitive vice.[2] If this default view is right, there is simply not much to be said about the epistemology of forgetting.

The default view flows from a simple, intuitively plausible characterization of the function of memory (what might be termed "preservationism about memory function"),[3] according to which memory's role is to preserve information acquired in the past, making it available again for future use. But this characterization of the function of memory is (as will emerge) at best a crude oversimplification. Thus the neglect of forgetting by epistemologists, though natural, is unfortunate, and the purpose of this paper is to begin to rectify it: I develop a detailed epistemology of forgetting, arguing, against the default view, that normal human forgetting—the pattern of forgetting characteristic of cognitively normal adult human beings—approximates a virtue located at the mean between the opposed cognitive vices of forgetting too much and remembering too much. In Sect. 2, I argue that, for any finite cognizer, a certain pattern of forgetting is necessary if her memory is to perform its function well. In Sect. 3, I argue that, by eliminating "clutter" from her memory store, this pattern of forgetting improves the epistemic properties of the subject's total doxastic state. I conclude, in Sect. 4, by reviewing work in psychology which suggests that normal human forgetting approximates this virtuous pattern of forgetting.

I take the default view to be the default not because it is often stated explicitly by epistemologists—it is not—but rather because, in addition to being suggested by the natural view of the function of memory, it is suggested both by the folk psychology on which many epistemologists continue to rely and by a certain unfortunate feature of the scientific psychology which others have adopted. Folk psychology tends to view all instances of forgetting as instances of memory failure: we routinely bemoan our "bad" memories, thinking of occasions on which we could not remember what we wanted to remember; when one cannot remember what she wants to remember, she is likely to say literally that her memory "fails" her. Even once we leave folk psychology behind, we tend today to rely on a computer model of memory: we have by now grown used to thinking of cognition (correctly) as computation, but this has had the unfortunate side-effect of leading us to think of the human mind on the model of an electronic computer and, in particular, to think of human memory on the model of computer memory; and as long as we take the computer model for granted, forgetting will appear to be obviously vicious—"forgetting", in the computer case, is indeed always memory failure.[4] (Folk psychology, of course, is a thoroughly unreliable guide to the actual workings of the mind. And computer

[2] Though it is convenient to express the default view in this way, some might object to the formulation on the ground that only traits which are defects relative to those typical of the relevant larger population can count as vices. If the account of cognitive virtue set out in section 2 below is right, however, the formulation is acceptable, since the account does not require that cognitive vices be atypical.

[3] Not to be confused with preservationism about memorial justification (Lackey 2005; Michaelian 2010a).

[4] The psychologist Gary Marcus has explicitly argued for something like the default view (2008); note that his discussion of memory and forgetting relies heavily on the computer model.

memory is an extremely poor model for human memory—human memory and computer memory are designed to solve different information-processing problems and therefore obey different principles.)

My argument assumes a definite account of cognitive virtues (set out in Sect. 2), and it might be worried that this limits its import significantly, that those who do not accept the account (or at least a fairly similar one) will have no reason to accept my thesis. There are two points to be made in response to this worry. First: The basic point that I want to make about normal human forgetting—that it is, contra the default view, not in general epistemically bad—is one that should be able to be stated and defended in non-virtue theoretic terms or in terms of a different virtue-theoretic framework. The virtue-theoretic framework adopted here is a convenient one in terms of which to develop the basic strategy of my argument, but it should be possible (even if more difficult) to develop analogous arguments without even using the concept of cognitive virtue. In other words: within any plausible epistemological framework, we should be able to recognize that forgetting is necessary for finite cognizers and that, given that forgetting is necessary, a certain pattern of forgetting is preferable. Second: The argument does not assume any particular connection between the virtues and knowledge or justification—it does not, in other words, depend on a full-blown virtue theory. I support the thesis that normal human forgetting is virtuous by pointing to certain features that are desirable in memory systems for cognizers like us and by then arguing that the human memory system comes reasonably close to having those features. Virtue-reliabilists who define knowledge/justification in terms of virtue by defining virtue in terms of reliability should still admit that the features in question are desirable (though they will maintain that they are unnecessary for virtue). Virtue-responsibilists, with their focus on intellectual character traits, can of course agree that there are better and worse ways for cognitive systems to perform their functions. And non-virtue epistemologists who define knowledge/justification in other terms similarly are free to accept my characterization of a memory system that performs its function well.

The remainder of Sect. 1 first precisifies my thesis and then attempts to raise it to the level of initial plausibility.

The restriction of the scope of the thesis to humans is necessary for two reasons. First: Forgetting is a process which unfolds within a memory system. I take it that a given cognitive system is virtuous to the extent that it performs its function well (Sosa 1991) and thus that a given pattern of forgetting is virtuous to the extent that it contributes to the performance of its function by the relevant memory system. Thus the thesis amounts to the claim that the pattern of forgetting characteristic of human beings is virtuous relative to the human memory system as it is in fact organized. This means that the fact that the same pattern of forgetting might not be virtuous relative to a memory system organized along sufficiently different lines poses no problem for my argument: e.g., the argument turns in part on the point that forgetting is a necessity for us given that our computational resources are limited (Sect. 2); the relativity of the virtuousness of normal human forgetting to the actual human memory system means that the claim that a creature with unlimited computational resources would not need to forget is irrelevant to my argument. Second: I take it that what it is for a system to be virtuous is for it to perform its function well not in any arbitrary environment but rather in the environment (or range of environments) that is normal for it (Sosa 1991). Thus the thesis amounts to the claim that normal human forgetting is virtuous relative to (the actual human memory system and) the range of environments typically encountered by humans. This means that the fact that the same pattern of forgetting might not be virtuous relative to a sufficiently different environment poses no problem for my argument: e.g., the argument turns in part on the point that normal human forgetting contributes to the performance of its function by memory in part because it in effect assumes that the environment is structured in a certain way (Sect. 4); the relativity of the virtuousness of normal human forgetting to the typical human environment means that the claim that this pattern of forgetting would not contribute to the achievement of its function by memory in an environment without this structure is irrelevant to my argument. The more specific restriction to cognitively normal adult human beings is meant to rule out several types of forgetting, including: the forgetting that occurs when the memory system suffers permanent damage (e.g., due to brain injury) or temporary interference (e.g., "alcoholic blackout"); the forgetting that accompanies the forms of cognitive decline associated with old age; and the forgetting of very young children ("childhood amnesia"). The epistemic status of the last of these types of forgetting is at best unclear; the others are clearly counternormative.

Given the restriction to normal adult human memory, there are further questions about which phenomena of memory the thesis is meant to cover. We might attempt to define forgetting by saying that it occurs when a subject once stored information but no longer does, but this definition lumps together phenomena that should be kept apart.

My focus here is on long-term memory (LTM), the system that stores stable representations both of general facts (semantic memory) and of our past experiences (episodic memory). Declarative memory is to be distinguished both from procedural memory and from working memory. Procedural memory arguably does not store information (Michaelian 2010c), and so procedural "forgetting" is likely an entirely distinct phenomenon. Working memory is a sort of limited-capacity mental workspace, and so loss of information from working memory is an utterly routine matter. Unlike working memory, short-term memory is not a distinct system: stable storage in LTM does not occur instantaneously (e.g., if little elaborative encoding occurs, then the record is unlikely to achieve stable storage); 'short-term memory' refers rather to the initial phase of storage of information in LTM. There is no consensus about how best to characterize the representations stored by LTM. I will use the general term 'record' to refer to these representations, whatever their form. I assume that some records are propositional, and it is these on which I focus.[5]

[5] The focus on propositional records allows me to bring existing epistemological frameworks into contact with the psychology of memory in a relatively straightforward fashion, but the basic strategy of the argument can be extended to cover other types of records as well, including imagistic records. While considerations similar to those to which I appeal to show that forgetting of propositional records is necessary for epistemic virtue also suggest that forgetting of imagistic records is necessary for virtue, the argument for the latter conclusion will involve additional complexities, since frameworks for assessing

The structure of long-term memory processes means that loss of information can occur at various stages: a representation might achieve storage in the form of a stable record in LTM but then be lost for some reason; a representation that survives in short-term memory might nevertheless not endure in the form of a stable record in LTM; and a representation might be stored temporarily in working memory without even a short-term record being encoded (in which case it will not result in a stable record in LTM). I group the latter two types of loss of information together under the heading 'non-encoding' (since, in the former type, while a short-term representation exists, no stable long-term record is encoded, and since, in the latter type, not even a temporary short-term record is encoded).[6] In ordinary language, 'forgetting' refers both to non-encoding and to loss of information from LTM. But there are important differences between the two types of loss of information. There are differences at the psychological level: the processes responsible for non-encoding and loss of information from LTM are distinct—non-encoding is a matter of failing to store information (more precisely: of storing it only temporarily), while loss of information from LTM is a matter of eliminating long-stored representations. And there are differences at the epistemological level: we will readily grant that non-encoding is often normative, whereas we tend to view loss of information from LTM as generally counternormative. (This is presumably because non-encoding is pervasive: the overwhelming majority of our beliefs are ephemeral—we hold them for a only short period of time before they vanish permanently.) In what follows, I reserve 'forgetting' for loss of information from LTM (though I return to non-encoding in Sect. 3).

Forgetting, then, is loss of information from LTM; but this is still not a satisfactory definition, for it says nothing about the nature of the loss in question. We tend ordinarily to think of forgetting as involving the (gradual) permanent elimination or erasure of records. But though information is sometimes permanently eliminated from memory, this is not the only sort of loss involved in forgetting; nor, in fact, is it even the primary sort of loss.

We can distinguish between forgetting in the sense of the permanent elimination of a memory trace (the unavailability of a record) and forgetting in the sense of the (possibly temporary) inaccessibility of a trace.[7] Once stored in LTM, relatively few traces are actually rendered unavailable: as Bjork and Vanderhuele point out, forgotten items "can typically be recognized at a rate that greatly exceeds chance levels, can be relearned at an accelerated rate, and can often be recalled in special

Footnote 5 continued
the epistemic adequacy of cognitive processes involving such non-propositional representations are less well-developed.

[6] There are psychological differences between the two types of non-encoding; see Schacter 2001 on transience vs. absent-mindedness. But (at least at the level of generality at which I am working here) they can be treated together for epistemological purposes.

[7] The inaccessibility/unavailability distinction was first drawn by Tulving and Pearlstone (1966). Note that inaccessibility is often subjectively indistinguishable from unavailability. (The "tip of the tongue" phenomenon, in which a subject feels that she knows something but is presently unable to recall it (Metcalfe 1994), is an exception.) This probably accounts for our tendency to assume that forgetting is normally a matter of the complete elimination of a record.

circumstances that reinstate certain cues from the past—all of which constitute evidence that such items have not been lost from memory in any absolute sense". The implication is that

> [o]nce information is successfully embedded within the knowledge network that defines long-term memory, it appears to remain in storage essentially forever. Even the most overlearned and heavily used items of information ... eventually become non-recallable with a long enough period of disuse, but such forgetting is a matter of loss of retrieval access to such items, not a loss of their representation in memory per se. (1992, p. 156)

In other words, most forgetting takes the form of inaccessibility rather than unavailability. In what follows, I set aside the relatively rare cases in which information is actually eliminated from LTM to focus on the inaccessibility of stored records: when I say that a subject has forgotten that *P*, I will generally mean that though her LTM stores a record that *P*, she cannot now retrieve it in response to appropriate stimuli (relevant stimuli that are available in her current environment).[8,9]

Having precisified the thesis, I turn now to the task of raising it to the level of initial plausibility. Following Aristotle, we generally think of a virtue as being situated at the "mean" along a continuum between two opposed vices. The basic thought behind the thesis that normal human forgetting is a cognitive virtue is that we can identify a pair of opposed cognitive vices of memory the mean between which defines a virtuous type of forgetting; normal human forgetting is virtuous because (and to the extent that) it approximates this mean.[10] Speaking loosely, we can say that the vices in question are those of forgetting too much and remembering too much, but the virtuousness or viciousness of a given pattern of forgetting should not be understood in purely quantitative terms—virtuous forgetting is a matter of forgetting the right records rather than simply of forgetting the right quantity of records.

Consider, e.g., a subject with retrograde amnesia, a condition (normally caused by brain injury) in which one suddenly loses access to a significant portion of her long-term memory store. Cases like this, in which the subject forgets significantly more than is normally forgotten, provide clear (though extreme) examples of

[8] This definition is less strict than that advocated by Wixted (2007), who proposes that we should say that forgetting occurs only if the cueing conditions in effect when retrieval is attempted are precisely the same as those that were in effect at an earlier time. While this strict definition might be appropriate for laboratory studies, where cueing conditions can be precisely reinstated, it has the consequence that forgetting can rarely be observed in non-laboratory contexts, where it is rare that precisely the same cues are used at different times.

[9] Note that the recognition that forgetting is a matter of inaccessibility rather than unavailability should already render us less reluctant to accept the thesis that normal human forgetting is virtuous, for if forgetting is a matter of records becoming inaccessible, then it is not irreversible. See Sect. 2.

[10] I emphasize that the vices in question are cognitive: there are cases in which forgetting might be prudentially but not cognitively appropriate (e.g., memories of personal trauma (Liao and Sandberg 2008)), cases in which forgetting might be cognitively but not morally appropriate (e.g., memories of one's ancestors (Blustein 2008)), etc. The interactions among cognitive, prudential, and moral virtues and vices of memory is an interesting topic, but one which I cannot take up here.

vicious forgetting. Note that the problem in amnesia is not simply that the amnesic subject has retrieval access to too few records: if the same brain injury which eliminated access to memories from a certain period prior to the injury were somehow simultaneously to restore access to a similar quantity of previously inaccessible memories from another period in the subject's life, it would still involve vicious forgetting—what is crucial is that the subject loses access to records that (intuitively) she needs to be able to continue to retrieve.

Though cases of remembering significantly more than is usually remembered are (less studied and therefore) less familiar, consideration of such cases begins to suggest that an unusually "good" memory, too, can be cognitively vicious.[11] "AJ",[12] studied extensively by Parker et al., provides the only known case of hyperthymestic syndrome, a condition in which the subject "spends an abnormally large amount of time thinking about his or her past" and "has an extraordinary capacity to recall specific items from their personal past" (2006, p. 47). Hyperthymesia differs from more familiar cases of extraordinary memory: the hyperthymestic's unusually "good" memory is confined to her personal past—she is not especially good at remembering arbitrary information in the manner of a mnemonist; and her unusual memory does not derive from the use of special mnemonic strategies of the sort used by mnemonists—hyperthymesia is a matter of having an unusual memory system rather than of using a normal memory system in an unusual way (2006, p. 36).[13]

It is important to note that AJ does not simply spend an unusual amount of time dwelling on her personal past but also has an unusually high degree of access to her autobiographical memories. She can recall far more about her personal past than can a normal subject: given a date within the period covered by her exceptional memory, she can provide detailed, specific information on what happened to her on that date; a normal subject does not even approach this degree of retrieval access. These two features combine to produce her unusual pattern of remembering and forgetting. Parker et al. suggest that AJ's unusually "good" memory can be explained in part by her inability to turn off "episodic retrieval mode": in effect, present stimuli always act as retrieval cues for her; thus, there is a sort of feedback loop in which one stimulus results in the retrieval of a large number of memories, which in turn are interpreted as retrieval cues, resulting in the retrieval of additional memories, and so on (2006, p. 39). But clearly her memory would be extraordinary even if she could turn off episodic retrieval mode: normal subjects simply do not have such extensive access to their episodic memories.

[11] Certain cases of persistence (in which a subject continues to recall a memory that she would rather forget) also seem to suggest that improved recall does not always result in an improvement to memory. See Schacter 2001 for discussion of persistence.

[12] After the article in which her case is discussed was published, AJ identified herself publicly as Jill Price.

[13] Though I rely here on examples of vicious remembering and forgetting due to abnormal memory systems, there can also be cases of vice due to abnormal uses of normal systems. (The well-known case of "S" (Solomon Shereshevsky), discussed by Luria (1987), might be of this type.) Given my focus on cognitive systems as virtuous or vicious—see Sect. 2—cases of the latter sort are less central here, though a complete treatment would have to take them into account (perhaps using a Zagzebski-style account (1996) of high-level virtues).

At first glance, AJ might appear to have an enviably good autobiographical memory. But closer examination of the case suggests that though we naturally assume that increased access to stored memories (less forgetting) would amount to an improvement to memory, this is not in fact the case. There are two points to note here. First: Though it is natural to assume that a "better" memory would provide us with a significant cognitive advantage, this is likely not the case. As Parker et al. point out, AJ's exceptional memory has provided her with no apparent advantage in daily life or in her studies; nor is it helpful on IQ tests and the like (2006, p. 48). And at the same time, AJ's unusual retrieval capacity carries heavy cognitive costs. In particular, she "spends much of her time recollecting the past instead of orienting to the present and future" (2006, p. 48). An increased retrieval capacity comes at a price: time that would otherwise be spent on other cognitive tasks is devoted to retrieval; time that would otherwise be spent acquiring new knowledge is spent simply processing "surplus" retrieved memories.

It might be thought that if AJ's exceptional retrieval capacity could be dissociated from her tendency to spend an unusual amount of time thinking about her past, then these costs could be minimized. But even if such a dissociation is feasible in principle, the resulting form of memory does not seem to amount to an improvement on normal human memory. This is the second point to note about the case: most of the surplus information that AJ can retrieve is, intuitively, trivial. She can, e.g., recall what she was doing on every Easter for most of her life (and even on what day Easter fell in a given year). It is at best unclear how a tendency to recall (along with useful information) great quantities of trivia could represent an improvement to memory. Significantly increased retrieval access would no doubt prevent us from forgetting some memories that we would like to retain; but it would also prevent us from forgetting a much greater quantity of memories that we are not interested in preserving.[14]

Consideration of AJ's case thus suggests that there is a vice of remembering too much opposed to the vice of forgetting too much: in the latter case, the memory system fails to preserve access to information that the subject still needs, while in the former case, the system continues to preserve access to information that the subject no longer needs. The existence of these opposed vices of memory, in turn, suggests the existence of a virtuous form of forgetting approximated by normal (non-amnesic, non-hyperthymestic) human memory: virtuous forgetting will be a matter of achieving the mean between these two extremes; (roughly) in virtuous

[14] Whether a given subject is interested in information of a given type surely depends in part on her ability to retrieve information of that type; thus if we had a greater capacity to recall certain types of trivia, the information in question might cease to be trivia for us. I assume that we should assess a given cognitive faculty relative to the actual interests of the relevant subject rather than relative to the interests that she would come to have, were she to have the given faculty; thus the dependence of our interests on our retrieval capacities does not threaten my claim that a tendency to recall great quantities of trivia would not constitute an improvement to memory. AJ herself continues to be interested in the surplus information that she is able to retrieve (Parker et al. 2006, p. 39), so, strictly speaking, it is not trivia for her (at least not in the sense in which I use the term in Sect. 3). But presumably she would not continue to be interested in the information if she were not constantly and automatically retrieving it. In other words: most of the surplus information that a normal subject would be able to retrieve if she had AJ's exceptional retrieval capacity is information in which she is not in fact interested and so, strictly speaking, is trivia.

forgetting, the subject forgets information that she no longer needs, retaining access to information that she continues to need.

2 Retrieval and the Finitary Predicament

I favour an account of cognitive virtue broadly similar to that incorporated in the sort of virtue-reliabilism associated with Sosa (1991), according to which virtues are certain cognitive faculties (which I take to include cognitive systems).[15] Though I take Sosa's conception of cognitive virtues—he describes them (roughly) as stable, reliable faculties—as my starting-point, I modify the conception by requiring properties in addition to reliability for virtue:[16] Sosa aims to provide an account of knowledge and therefore takes reliability to be central to virtue; but, as Goldman has pointed out, reliability is only one among a number of epistemically important properties of cognitive processes or systems—though these properties are not reflected in the concept of knowledge, we care also about power and speed, in particular (1992).

A system or process is reliable if it tends to produce mostly true beliefs, that is, if the ratio of cases in which it produces true beliefs to cases in which it produces beliefs (whether true or false) is high. Thus the reliability of a system is compatible with its not producing very many true beliefs, and a system that produces too few true beliefs will be epistemically deficient—we care not only about the reliability of a system but also about its power, about whether it tends to produce many true beliefs. Goldman defines the power of a system or process as the ratio of cases in which it produces true beliefs to cases in which it produces beliefs (whether true or false) and cases in which it fails to produce a belief. (Call this "power-1".) But a given process-token might produce multiple beliefs—a single act of retrieval, e.g., can produce many memory-beliefs—and so we are interested also in the sheer quantity of true beliefs that a system or process is capable of producing ("power-2"). Note that though more reliability is (ceteris paribus) presumably always better, plausibly it is not the case that more power is always better; if, e.g., the additional beliefs that would be produced due to an increase in the power of a system are trivial, the increase might not amount to an improvement to the system. Reliability and power are necessary for virtue, but they are not sufficient: a system that is highly reliable and highly powerful might be slow—given a choice between two equally reliable and powerful systems, one of which produces its outputs more quickly than the other, we will prefer the faster system. Note that while there are sometimes trade-offs between reliability and speed, and between power-2 and speed—one system might be faster than another, but at the cost of some reliability;

[15] This conception of virtue is particularly well-suited to my purposes here, as my focus is on the evaluation of a specific cognitive system. But, though Sosa's conception contrasts with the virtue-responsibilist conception of Zagzebski (Zagzebski 1996), according to which virtues are certain acquired intellectual character traits, the virtue-responsibilist should be able to accept my basic conclusions (though she will want to rephrase them in other terms).

[16] Lepock (2009) develops a similar approach in more detail; note that whereas I require speed for virtue, he replaces speed with the more general property of portability.

increasing the power-2 of a system will tend to slow it down—speed is in general a prerequisite for power-1: if a system is too slow, its processes will (at least if the total cognitive system of which it is a component is well-designed) in many cases terminate without outputting beliefs, as they are interrupted to divert resources to other tasks.

I will therefore take a virtuous cognitive system to be one which achieves appropriate levels of reliability, power, and speed.[17] The levels that are appropriate for a given system are determined by its function: in general, a virtuous system is one that performs its function well; the appropriate levels are those necessary for the system to perform its function well.[18] Sosa in effect takes the function of any cognitive system to be getting at the truth, but this is an overly general view of the functions of cognitive systems. I know of no explicit recipe for determining the function of a given cognitive system, but I do suppose that we should avoid assuming at the outset of our investigation of the system that we know its function; in order to determine its function, we should rather look to its role in the complex of systems of which it is a part, in the cognitive life of the subject. Given this general conception of cognitive virtue, a virtuous memory system will be one which achieves appropriate levels of reliability, power, and speed, where the appropriate levels are determined by the function of memory, which function is in turn determined by the role of memory in the cognitive life of the subject.[19]

Given this account of cognitive virtue, what can be said in favour of the thesis that forgetting is necessary for virtuous memory? Cherniak emphasizes that a basic feature of the human situation (and, indeed, of the situation of all but the most absurdly idealized cognizers) is that

> human beings are in the *finitary predicament* of having fixed limits on their cognitive capacities and the time available to them. Unlike Turing machines,

[17] One might suspect that I have selected this nonstandard account of cognitive virtue precisely because it generates my desired conclusion that forgetting is necessary for virtuous memory. But, first, the account is independently plausible—power and speed are plainly desirable features in cognitive systems. Second, more standard accounts of virtue make it difficult even to ask the question about the normative status of forgetting, simply because they have nothing to say about the potential contributions of processes which eliminate but do not produce beliefs, and this is a legitimate reason for favouring my account. Finally, the account does not in fact by itself imply that normal human forgetting is virtuous but only that a certain pattern of forgetting is virtuous for creatures with finite computational resources; the extent to which normal human forgetting approximates this pattern is a further, empirical question.

[18] It is plausible that the function of the system determines only a range of permissible levels, which implies that systems of the same type which employ somewhat different balances of reliability, power, and speed can all qualify as virtuous. My argument is compatible with this possibility, since it aims only to show that forgetting is necessary for the attainment of appropriate levels of reliability, power, and speed, without specifying a precise balance of these properties.

[19] A question arises at this point about whether the reliability and power of memory are to be understood as conditional or as unconditional, belief-independent or belief-dependent. It will make sense to treat memory as belief-dependent if a record that *P* is normally stored in such a way that the subject is disposed to accept it when it is retrieved as a consequence of the subject's accepting the content that *P* (believing that *P*) at the time of the encoding of the record, so that we can treat the earlier belief that *P* as an input to memory (despite the fact that memory does not literally store beliefs). I will, however, assume that the subject's other cognitive systems are largely reliable, so that the majority of beliefs given to memory as inputs are true, which means that I will in general be able to ignore the distinction between conditional and unconditional reliability in what follows.

> actual human beings in everyday situations or even in scientific inquiry do not have potentially infinite memory and computing time. (1986, p. 8)

The finitary predicament plays a crucial role in establishing the necessity of forgetting for virtuous memory; but care is required in making this argument.

The finitary predicament has two aspects: first, the human memory system has finite storage capacity; second, human cognition takes time (i.e., we can only perform so many computations in a given length of time). Harman in effect appeals (in part) to the first aspect of the finitary predicament in order to argue for a principle of "clutter avoidance" in belief-updating:

> There is a limit to what one can remember, a limit to the number of things one can put into long-term storage, and a limit to what one can retrieve. It is important to save room for important things and not clutter one's mind with a lot of unimportant matters. (1986, pp. 41–42)

We might similarly attempt to argue that forgetting is a cognitive virtue by appealing to the limited storage capacity of human memory: if memory has a finite storage capacity, perhaps some old records must be forgotten in order to make room for incoming records (though this would leave the further question of which records should be forgotten). But this is not a feasible line of argument, for though the capacity of LTM is indeed finite, it is unlimited in practical terms.

Because her storage capacity is finite, if a human being were to live for a sufficiently long time, she would eventually run out of capacity. But in fact we are not in danger of running out of capacity. Nor is this because we forget as much as we do: we would not run out of capacity even if we forgot significantly less than we do, for most forgetting takes the form of inaccessibility (rather than unavailability)—most forgotten records are still stored in memory. The practically unlimited capacity of human memory derives rather from its structure. One way in which the (unfortunately standard) computer model of memory is misleading is that the capacity of a computer memory is fixed, whereas that of a human memory is not: the addition of a record to a computer memory uses up some of the memory's fixed storage capacity, but the addition of a record to a human memory can actually create new capacity.[20] Thus the suggested argument for the necessity of forgetting depends on an untenable assumption: there is no interesting limit on the amount of information that we can hold in long-term storage; and hence an appeal to our finite storage capacity will not establish the necessity of forgetting for humans.

It is possible to challenge the claim that forgetting is not rendered necessary by finite storage capacity on the basis of certain features of connectionist networks, in which catastrophic forgetting can occur due to the superpositional storage of similar items (McCloskey and Cohen 1989; Ratcliff 1990)—arguably, in these networks, finite storage capacity requires forgetting. However, this type of forgetting does not

[20] Bjork and Bjork use the metaphor of scaffolding to illustrate this feature of human memory: "We are fond of telling laypersons that our memories are not like a box in the sense that storing some information leaves less room for additional information. Rather, we say, a more appropriate analogy is that our memory is like a scaffolding structure of some kind such that the more developed (or elaborated) the structure the more additional ways there are to enter (or attach) new information" (1988, p. 285).

normally occur in biological memory systems (including the human memory system) (French 1999).[21] So while it might not be true in general that forgetting is not rendered necessary by finite storage capacity, the claim holds when restricted to human memory, which is my target here.

While the necessity of forgetting cannot be established by appealing to finite storage capacity, forgetting is indeed rendered necessary by the second aspect of the finitary predicament, limited computational resources.[22] Bjork and Vanderhuele argue that "[i]f we take as a starting point that humans are remarkable as storage devices, and that there are obvious advantages of having virtually unlimited capacity in that domain, the limitations on retrieval access can be viewed as a necessary filter. In the interest of speed, accuracy, and avoiding confusion, we do not *want* every item in our memories to be accessible" (1992, p. 157). My approach here can be viewed as a development of this suggestion: limitations on our computational capacities mean that, without forgetting, memory would not be able to perform its function well, that forgetting is necessary to ensure appropriate levels of reliability, power, and speed in memory.[23]

However, precisely, the function of memory is to be characterized, a high level of reliability will be necessary for the adequate performance of that function: it is a drastic oversimplification to characterize the function of memory simply as that of making information acquired in the past available again for current use; but its function clearly involves some sort of making-available-again, and a high level of reliability is necessary for the adequate performance of any such function.

I assume the following general picture of retrieval.[24] The records stored in memory are organized so that the relevance of a given record to a given query can be determined. Retrieval occurs when a query is sent to memory. All records that are both relevant to the query and accessible are retrieved:[25] if a record is irrelevant, it is not retrieved; if it is relevant but inaccessible, it is likewise not retrieved. The

[21] See McClelland, McNaughton, and O'Reilly (1995) for an influential explanation of this fact.

[22] See the discussion of the relativity of the virtuousness of forgetting to the actual human memory system in Sect. 1 above.

[23] It might be worried at this point that, depending on how the details of the default view are spelled out, the difference between that view and the view that I defend largely disappears. It we take the defender of the default view to be claiming that it is unfortunate that we have the finite computational resources which (I argue) render it necessary that we forget, the difference becomes one merely of emphasis: whereas the default theorist emphasis that it is unfortunate that we are in a situation that renders forgetting necessary, I emphasize that, given that we are in such a situation, it is fortunate that we forget. But the default view should not be understood this way. First, the default view on the epistemology of forgetting is plausibly taken precisely not to say anything about the implications of the fact of our finite computational resources—though there are a growing number of exceptions, most epistemological theories have not seriously taken the fact of finite computational resources into account. Second, a view which says that it is unfortunate that we have finite computational resources would anyway be strange, for the computational resources of any physical cognizer necessarily have epistemologically significant limitations.

[24] I simplify by treating records as if they were discrete, an assumption challenged by views on which memories are stored only in a distributed, superpositional manner (Sutton 1998). A more realistic picture of the nature of memory traces would complicate my argument but should not affect the success of its basic strategy.

[25] This is a simplifying assumption. See the discussion below of Cherniak's description of the memory store as compartmentalized for a more precise statement.

outcome of successful retrieval is an occurrent belief or beliefs.[26] In some cases, some or all retrieved records must be discarded before the retrieval process concludes; e.g., if some of the records retrieved in response to the query “my telephone number” are for numbers that I no longer have (and if I realize this), these will normally be discarded as part of the process of forming a belief that my telephone number is *X* (or a set of similar beliefs). If no relevant record is retrieved (or if all retrieved records are discarded), no occurrent belief is produced. If this general picture is right, retrieval is computationally costly: search takes time—if more records are tested for relevance, search takes longer; and sorting through accessible relevant records to determine which ones are wanted takes time—if more relevant records are identified, sorting takes longer. Moreover, the immediate consequences of retrieval are computationally costly: retrieved records will often trigger additional thoughts—retrieving more records generally requires more additional thinking.

Reliability is essentially a matter of avoiding the formation of false beliefs; in the context of memory, this amounts to avoiding the formation of false occurrent beliefs as a consequence of the retrieval of inaccurate records from memory. If forgetting were to eliminate more accurate records than inaccurate records (more precisely: accurate or inaccurate records that the subject is disposed to accept),[27] it would diminish the reliability of memory. Thus a virtuous memory system will not tend to forget more accurate records than inaccurate records. By the same token, if forgetting were to eliminate more inaccurate records than accurate records (accurate or inaccurate records that the subject is disposed to accept), it would (by reducing the frequency of cases in which the subject retrieves an inaccurate record and consequently forms a false occurrent belief) increase the reliability of memory. Thus a virtuous memory system might incorporate a certain pattern of forgetting as a means of increasing the reliability of memory. Obviously, the memory system has no way of directly determining whether a given record is inaccurate, and so it cannot target inaccurate records as such. But it is nonetheless possible for it to forget inaccurate records preferentially (to forget them at a higher rate than accurate records).

I assume that we are concerned with subjects whose other cognitive systems are largely reliable, so that the records stored in their memories are mostly

[26] It is natural to think of memory as storing and retrieving beliefs, but the thought involves a confusion. I take it that a subject has an occurrent belief that *P* when she has an activated representation that *P* that plays a certain role in her mental life—roughly: she accepts the representation as true (This is crude, but subtle differences among different conceptions of belief will not affect my argument here.). This can occur when a record is retrieved from LTM to working memory; but the record stored in LTM (obviously) is not an occurrent belief. I take it that a subject has a dispositional belief that *P* when she has a record that *P* stored in her LTM, she is disposed to retrieve the record in response to relevant stimuli, and she is disposed to form an occurrent belief with the record as its content (to accept the record as true) if the record is retrieved. (Not every record stored by memory would be believed if it were retrieved. Memory stores records stemming from imagining, fantasizing, etc. And memory normally continues to store the record that *P* even after the subject has abandoned her belief that *P*.) Though the subject will have a dispositional belief that *P* in part because she stores a record that *P* in LTM, long-term memory does not actually store the dispositional belief.

[27] Memory also stores records of dreams, fantasies, etc., which the subject will normally not be disposed to accept. I bracket these in what follows.

accurate;[28] but this does not mean that forgetting cannot make a significant contribution to the reliability of memory, for even if a record is accurate when initially stored, it need not remain so.[29] We can think of this as the problem of outdated information: the world changes around the subject, and records that once were accurate become inaccurate; thus even if most of a subject's records are accurate when initially stored, many of them will cease to be so over time.[30] Now, if there is no way for the memory system directly to determine whether a given record is inaccurate, there is no way for it directly to determine whether a given record is outdated, and so it cannot target outdated records as such. But if subject is rational overall, it is nonetheless possible for the memory system preferentially to forget outdated records and thus preferentially to forget inaccurate records.

A rational subject will on average tend to retrieve recently-acquired records more often than older records, simply because her environment is more likely to provide retrieval cues for recently-acquired records: as the subject moves from one environment to another, or as the environment around her evolves, many older records lose their relevance to her current situation.[31] And the older a record is, the greater is the chance that it has become outdated: as time passes, the probability that the world has invalidated a given record increases. Thus the records that are retrieved less often by a rational subject will disproportionately include outdated records. This means that if the memory system is sensitive to the retrieval history of records, it can preferentially forget outdated records: if the system tends to render records inaccessible when they are infrequently retrieved, outdated records will be rendered inaccessible at a higher rate than current records. Thus the memory system can in principle preferentially forget inaccurate records, and thus forgetting can in principle increase the reliability of memory.[32]

The importance of this potential contribution of forgetting to the reliability of memory should not be underestimated. Even where the subject actively updates her beliefs, the record that underwrote an outdated belief can continue to lower the reliability of retrieval. The following sort of scenario is common: A subject believes

[28] It might be objected that I am not entitled to this assumption. This is not the place for the defence of such a general assumption, but I note that similar assumptions are made by many theorists; e.g., though they are interested specifically in misbelief, McKay and Dennett cite, in addition to Dennett's work (1987), that of Fodor (1983) and Millikan (1984), as assuming that humans "have been biologically engineered to form true beliefs—by evolution" (2009, p. 493). It might be objected that the assumption starts to look particularly problematic when we consider the extent of our reliance on testimony; for a defence of the claim that formation of testimonial belief can be reliable despite our vulnerability to deception, see Michaelian 2010b.

[29] The complete story about the reliability of memory will need to explain how memory can be reliable despite its constructive character; see Michaelian 2010a for an explanation of the compatibility of construction and reliability, emphasizing the role of metamemory.

[30] Note that I am here using 'outdated' in a narrower sense than it often has in discussions of forgetting: the term often refers to information that is no longer relevant to the subject's interests, whether or not it is still accurate.

[31] The point is statistical: obviously, certain features of the environment are more or less fixed or at least invariant over long periods of time, so that not all records lose their relevance; the subject will continue to retrieve these records regularly.

[32] The suggestion that forgetting is sensitive to retrieval history is not ad hoc; as we will see in Sect. 4, forgetting is indeed governed in part by retrieval history.

that *P* and stores a record to that effect in her long term memory in such a manner that she has a dispositional belief that *P*. She now learns that *Q*, recognizes that *Q* is incompatible with *P*, abandons her belief that *P*, and stores a record to the effect that *Q* in her long term memory in such a manner that she has a dispositional belief that *Q*. Though she no longer has a dispositional belief that *P*, the record that *P* does not vanish from her memory. Thus it will initially be retrieved (along with the record that *Q*) by relevant queries. When it is, there is a chance that the subject will end up forming an occurrent belief that *P*; this can happen, e.g., if her attention is divided. Thus the continued accessibility of the record lowers the reliability of memory. But though the record that *P* and the record that *Q* will be retrieved in response to some of the same queries, they will normally also be retrieved in response to different queries. This means that if the subject sends queries relevant to *P* to her memory less often than she sends queries relevant to *Q*, the former record will (if forgetting is governed by retrieval history in the manner suggested above) eventually become inaccessible, thus decreasing the probability that the subject will form a false occurrent belief.

It might be objected that a more virtuous memory system would achieve even greater reliability by eliminating the record that *P* entirely as soon as the incompatible record that *Q* is acquired: in such a system, the record that *P* would be deleted or overwritten with the new record that *Q*, thus eliminating the possibility of retrieving the record that *P* (and thereby the risk of mistakenly accepting the record as true). But note, first, that this approach provides only a partial solution to the problem of outdated information, since information often becomes outdated without the subject learning that it has become outdated. Note, second, that the solution is not feasible: instantaneous deletion or overwriting is not a realistic possibility for biological memory systems. Note, finally, that even if were feasible, it would not be desirable: records that have been invalidated might at some point become valid again; in such cases, it is beneficial to be able simply to restore access to the formerly outdated records rather than having to encode them again from scratch. This is an instance of a more general point: access to a record that has been forgotten (whether or not it is outdated) can be restored more quickly if forgetting takes the form of inaccessibility than it can if forgetting takes the form of unavailability (if records are overwritten or deleted); since the memory system cannot predict the future with certainty, a record that has been forgotten might be needed again in the future, and thus loss of retrieval access is preferable to deletion.[33]

Of course, if records are targeted for forgetting on the basis of their retrieval history, forgetting will eliminate access not only to outdated records but to any record that is seldom retrieved, whether or not it is accurate. Thus while forgetting might improve the reliability of retrieval (by preferentially eliminating access to outdated records), it will also reduce the power of retrieval: by reducing the number of records relevant to a given query that remain accessible, forgetting reduces the frequency of occasions on which retrieval will produce an occurrent true belief (the power-1 of memory) and the number of occurrent true beliefs that will be produced

[33] Kraemer and Golding (1997) and Bjork (1989) develop similar arguments.

by a given act of retrieval (the power-2 of memory). Thus it seems that there is a trade-off between reliability and power in retrieval.

But the trade-off is largely apparent, for it is a mistake to assume that more power is always better in retrieval. Assuming that the subject stores mostly accurate records, minimizing forgetting in a memory system will increase its power, since this will increase the number of occurrent true beliefs formed as a consequence of retrieval. But (as pointed out above) retrieval has computational costs (searching for relevant records and sorting through records identified as relevant to determine which are wanted) and computationally costly consequences (thinking occasioned by retrieved records); and the more powerful we make retrieval, the more extreme the costs and consequences of retrieval become. This means that while a certain level of power-1 in retrieval is obviously necessary—we do not want a memory system in which successful retrieval is a rare occurrence—increasing power-2 beyond a certain point has undesirable effects.

If retrieval were to produce too many records, memory would be worse than useless for the subject—she would be swamped by the beliefs output by retrieval. Absent forgetting, a given query might retrieve a huge quantity of records. This would make for a more powerful memory system: if you did not forget, sending a query to your memory would result in the production of a much larger number of true beliefs than you in fact obtain when you send the query. But such a system would be too powerful-2, for the computational costs and consequences of sending the query to memory would often be enormous: retrieval would take vastly longer than it does in fact; and coping with the resulting beliefs would be cognitively much more demanding than it is in fact.[34] Considering the role of memory in the broader cognitive life of a subject with limited computational resources thus allows us to refine our characterization of the function of memory: the function of memory is to make information acquired in the past available again for current use, but only in manageable quantities. Given this more refined characterization, we can see that while high levels of reliability and power-1 are appropriate for memory, only a more modest level of power-2 is appropriate. And forgetting is necessary in order to lower the power-2 of memory to this level.

While forgetting is necessary for lowering the power-2 of memory to an appropriate level, it is also necessary for raising the power-1 of memory to an appropriate level. This follows from the necessity of forgetting for an acceptable level of speed in retrieval. In general, improving speed improves power-1 (assuming that the system is reliable), since processes in a faster system will less often be interrupted in order to divert cognitive resources to other tasks before they can run to conclusion; if retrieval were too slow, it would be insufficiently powerful-1. I noted above that retrieval is computationally costly in two ways: searching through accessible records for relevant records takes time; sorting through the records that have been identified as relevant to determine which are currently wanted takes additional time. Forgetting improves the speed of retrieval by limiting the number of

[34] Recall the case of AJ: AJ's memory is highly reliable, but it is remarkable especially in that it is exceptionally powerful: she is somehow able to retrieve far more records than those with normal memory systems can retrieve. This exceptionally powerful retrieval capacity comes at a cost: AJ spends a great deal of time simply processing her retrieved memories.

accessible records: as forgetting increases, fewer records are searched, increasing the speed of search; and the number of records judged to be relevant decreases, so that the time necessary for sorting decreases. Retrieval in a memory system in which little or nothing is forgotten would become ever slower as new records are added; it would quickly become extremely slow. Unless the subject's total cognitive system is badly designed (so that a retrieval process is allowed to run to conclusion no matter how long it takes), retrieval in such a system would often be interrupted to divert resources to more pressing tasks, resulting in a system with a low level of power-1.

Speed is necessary in memory not only because it is a prerequisite for a high level of power-1 but also because retrieval is time-sensitive (in the sense that taking too long to retrieve a record is just as bad as failing to retrieve it). The point can be put in terms of the function of memory: if retrieval takes too long, memory is useless for the subject—by the time the retrieval process outputs a belief (assuming that it is allowed to run to conclusion), the belief will no longer be of use to the subject. In other words, a second qualification must be built into the characterization of the function of memory: the function of memory is to make information acquired in the past available again for current use (1) in manageable quantities and (2) in a timely manner. Absent forgetting, retrieval would be too slow to allow memory to perform this function well. Thus some forgetting is necessary for virtuous memory.

It might be objected that forgetting is not really necessary for ensuring an appropriate level of speed in retrieval. Cherniak emphasizes that the memory store is structured or compartmentalized—he uses the metaphor of a hierarchically-organized filesystem in which records are contained in files which can also contain further files—and that this compartmentalization improves the efficiency of search by rendering it unnecessary to search the entire memory store in response to a given query (1986). One might suggest that if the compartments are sufficiently narrow, then forgetting will be unnecessary for ensuring an acceptable level of speed in retrieval. But the suggestion does not work. First (as Cherniak notes), if the compartments are too narrow, then the system will often attempt retrieval from an incorrect compartment—in other words: the compartments can only be narrowed so much, for while narrowing the compartments increases the speed of retrieval, it also decreases the power-1 of retrieval. Second (and more importantly), even if the compartments are narrowed radically, a given compartment can grow indefinitely large as time goes on if there is no forgetting. Thus the compartmentalization of the memory store is insufficient on its own to ensure an appropriate level of speed in retrieval—some forgetting is indeed necessary.

I conclude that a virtuous memory system for any subject with finite computational resources will inevitably involve some forgetting: in view of the limitations on the subject's computational resources, a certain amount of forgetting is necessary if her memory system is to make information acquired in the past available again in usable quantities and in a timely manner. But this does not yet give us a picture of the specific pattern of forgetting characteristic of a virtuous memory system, it does not yet tell us which records a virtuous system will forget.

We have seen that a virtuous memory system will be only modestly powerful-2, that is, that it will retrieve only a modest number of records in response to a given query. Suppose that we have a memory system which achieves this modest level of power-2 and which achieves appropriately high levels of reliability, power-1, and speed. The system might, for all that we have said so far, systematically provide information that is irrelevant to the subject's interests—the modest number of records retrieved on a given occasion might typically be irrelevant to the subject's interests. Simplifying by ignoring the compartmentalization of the memory store, a record is retrieved in response to a given query if it is both accessible and relevant to the query. A record can be relevant to a query sent by the subject without being relevant to the subject's current interests. But if the records retrieved by the system are normally irrelevant to the subject's current interests, the system is clearly defective. Thus I now suggest a third qualification to the description of the function of memory: the function of memory is to make (1) currently-relevant information acquired in the past available again for present use (2) in manageable quantities and (3) in a timely manner. If this modified characterization of the function of memory is right, a virtuous memory system will preferentially forget (render inaccessible) records that are irrelevant to the subject's interests.[35]

In order for this suggestion to have any real content, something needs to be said about the nature of interests. In the course of his discussion of "veritistic value", Goldman distinguishes among three types of interest, three senses in which a belief might answer a question in which a subject is interested (1999, p. 95): the belief might answer a question that the subject actively finds interesting, in the sense that she is subjectively curious about the answer to the question; the belief might answer a question in which the subject is disposed to be interested, in the sense that, if she thought about the question, she would actively find it interesting; and the belief might answer a question in which the subject should be interested, in the sense that it is in her objective interest to know the answer to the question. (The distinction among these three types of interest matters for the argument of Sect. 4 below; I refer to them there as "subjective interest", "dispositional interest", and "objective interest", respectively.) Goldman plausibly suggests that each of these three types of interest is sufficient for epistemic value, that is, that if the subject is interested in a question in one of these senses, then if she has a belief that answers it, the belief has epistemic value. Without taking a stand here on the connection between virtue and value, I propose that a virtuous memory system will forget records according to whether they are of interest to the subject in one of the senses identified by Goldman.

If this is right, then an ideally virtuous memory system would forget records when they cease to be of interest to the subject and no sooner. The specific computational limitations of a given type of cognizer, however, might mean that a subject of that type is bound to forget more records than this; in such cases, the best solution is for the memory system to prefer to forget uninteresting records (to render them inaccessible before rendering any interesting records inaccessible); assuming

[35] Additional modifications might be necessary to take into account the role of episodic memory in "mental time travel" (Boyer 2009; Tulving 1999; Suddendorf and Corballis 2007).

that interests come in degrees, a virtuous system will prefer to forget less interesting records before more interesting records.[36]

There remains the question of how a memory system could be engineered to forget uninteresting records preferentially; I return to this question in Sect. 4 below, but I have already hinted at the relevant mechanism: if forgetting is governed by retrieval history, and if the subject's interest in a record is reflected in the record's retrieval history, then forgetting can be sensitive to the subject's interests.

3 Storage and Clutter Elimination

I have so far considered forgetting from the point of view of retrieval only; in this section, I consider it from the point of view of storage, conceiving of forgetting not as a means of ensuring that memory performs its function well but rather as a means of improving the shape of the subject's total doxastic state (her total belief set).

In Sect. 1, I noted that there are important differences at the psychological level between non-encoding and forgetting: non-encoding is a matter of failing to encode an enduring record, whereas forgetting is a matter of losing retrieval access to an existing record. I said that there are also differences between non-encoding and forgetting at the epistemological level, but I want now to argue that these differences are fairly superficial: there is in the end a fairly tight analogy between the epistemology of non-encoding and the epistemology of forgetting.

As noted in Sect. 2, Harman appeals to our limited storage capacity (the first aspect of the finitary predicament) to argue for the principle of clutter avoidance (CA): "One should not clutter one's mind with trivialities", where trivialities are "matters in which one has no interest" (1986, p. 55). The principle seems basically to concern encoding: the central idea, I take it, is that one should encode a record only if one is interested in it. This attempt to ground CA by appealing to our limited storage capacity fails, for (as we saw in Sect. 2) the storage capacity of LTM is for practical purposes effectively unbounded: even if, in violation of CA, a subject were to encode memories indiscriminately, she would in practice not run out of storage capacity. But nevertheless the principle is a plausible one, and I think that we can

[36] I have argued that a certain pattern of forgetting is virtuous in the sense that it is crucial for the performance of its function by memory. But forgetting can have additional cognitive advantages: given the pattern of forgetting associated with a virtuous memory system, other systems can exploit forgetting, for forgetting can itself serve as a source of information. Schooler and Hertwig have examined the benefits of forgetting for certain "fast and frugal" heuristics, including the recognition heuristic (Gigerenzer et al. 1999), the rule (in the case of a two-alternative choice) that "[i]f one of the two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion" (2005, p. 611). This simple rule is a useful heuristic "because lack of knowledge is often systematic rather than random", so that "failure to recognize something may be informative" (2005, p. 611); the idea (counterintuitively) is that in certain contexts partial ignorance is better not only than total ignorance but even than lack of ignorance. Schooler and Hertwig suggest that forgetting can be beneficial because it enables us to use the recognition heuristic where we would otherwise be unable to use it: drawing on Anderson's rational analysis of memory (Anderson 1990), they argue that forgetting enhances the performance of the recognition heuristic; the core idea is that if forgetting is a function of the frequency and recency with which information is encountered, then it can benefit cognition by making the recognition heuristic available for use.

offer a better rationale for it.[37] I appealed above to the notion of interests to suggest an answer to the question of which records a properly functioning memory system will prefer to forget; I want now to appeal to the same notion to motivate CA. The idea is that rather than appealing to practical constraints on the subject's cognitive capacities to motivate the principle, we should appeal to the effects of obeying the principle on the shape of her total doxastic state.

A total doxastic state has various epistemically-significant properties. There can be no uncontroversial list of these properties, but some sort of coherence, e.g., will certainly have a place on any acceptable list—all will agree that increasing the coherence of a state tends to improve it epistemically. It is natural to say that the size of the state (the number of beliefs it contains) also matters to us: within certain limits, increasing the size of a state will improve it epistemically. But ultimately it is not the size of the state as such that matters to us: increasing the size of the state will not improve it if the beliefs added are not of interest to the subject. What we care about is rather the size of the set of interesting beliefs included in the state: increasing the size of that set improves the state epistemically.[38] This is not yet enough to ground CA: the principle prohibits the addition of clutter (uninteresting beliefs), but the epistemic impropriety of adding clutter does not follow from the fact that adding clutter does not improve the state epistemically—that fact implies only that adding clutter is not epistemically required. I suggest that we care also about the ratio of interesting beliefs to total beliefs (interesting beliefs plus clutter): if two total doxastic states contain the same number of interesting beliefs and one contains in addition some clutter (beliefs that answer no question about which the relevant subject is actively curious, no question about which she would be curious if she thought about it, no question about which it is in her interest to be curious), then the larger state is epistemically inferior to the smaller state. The idea is that an increase in the size of a subject's belief set need not improve the belief set epistemically, that, moreover, if the increase in size decreases the ratio of interesting to total beliefs, it will worsen the set epistemically. If this suggestion is right, then so is CA: if we prefer uncluttered belief sets to cluttered belief sets, then a subject who obeys CA will have an epistemically better total doxastic state than will an otherwise similar subject who does not obey the principle, even though her total state will include fewer beliefs. (Note that this point is independent not only of considerations of the subject's computational limitations but also of considerations of the function of reasoning.)

Most of one's beliefs at a given time are merely dispositional (i.e., dispositional and non-occurrent). I assume that one has a dispositional belief that *P* if one stores a record that *P*, one is disposed to retrieve the record that *P* in response to appropriate

[37] Harman's argument for CA does claim that there is "a limit to what one can retrieve" (1986, p. 41), which suggests that we might attempt to establish CA by an appeal to the second aspect of the finitary predicament. The idea would be that retrieval will not go well if one encodes clutter, presumably because then retrieval will often produce trivial beliefs. But the argument is not promising: given that forgetting is sensitive to retrieval history in the manner suggested in Sect. 4, forgetting will render trivial records inaccessible.

[38] Of course, we care also about whether the beliefs are true; here, as throughout, I assume that we are dealing with subjects who acquire mostly true beliefs.

stimuli, and one is disposed to accept the record if it is retrieved. Not every record stored in memory corresponds to a dispositional belief, for one might lack the disposition to accept a record if it is retrieved. Nor does every record stored in memory that one would accept if it were retrieved correspond to a dispositional belief, for one might lack the disposition to retrieve a record in response to appropriate stimuli—if the record is inaccessible, then one lacks the disposition to retrieve it. Thus forgetting, even though it does not (typically) strictly eliminate records from memory, and even though it does not directly affect which records one would accept if they were retrieved, can have an effect on the subject's total doxastic state: by eliminating access to stored records, forgetting eliminates the corresponding dispositional beliefs (if any).

If something like my suggested rationale for the principle of clutter avoidance for encoding—one should not clutter one's mind with trivialities—is right, then we must also accept an analogous principle of clutter elimination (CE) for storage:[39] if one's mind is cluttered with trivialities, one should remove them. If the epistemic quality of a total doxastic state (a total belief set) is reduced by adding clutter (uninteresting beliefs) to it, then by the same token it is improved by removing clutter from it—clutter elimination improves the ratio of interesting to total beliefs. Thus if forgetting tends to eliminate records corresponding to uninteresting beliefs (as I have argued that virtuous forgetting does), then it is licensed by CE.

Note that there is a role for clutter elimination even in the cognitive lives of subjects who obey CA. First: a subject who obeys CA need not do so perfectly—she might inadvertently end up believing some clutter, which should then be eliminated. Second (and more importantly): just as the accuracy of a record changes over time as the world around the subject changes, the status of a belief as (non-)clutter changes over time as the subject's interests change (whether as a function of changes in her environment or due to her intellectual development). The following sort of scenario is common: A subject is interested in whether *P* is true, comes to believe that *P*, and stores a record that *P* in such a way that she continues to have a dispositional belief that *P*. But over time, her interests change in such a way that her belief that *P* turns into clutter. According to CE, her belief that *P* should then be eliminated.

Note that this argument is independent of considerations of the function of memory: those who do not accept my characterization of the function of memory or who are unpersuaded by my argument from that characterization to the existence of a virtuous form of forgetting can accept the present argument. My claim is that there is a coincidence between the pattern of forgetting associated with virtuous memory and the requirements of CE: CE implies that a belief should be eliminated when it no longer interests the subject; virtuous forgetting, I have argued, tends to eliminate records that do not interest the subject, thereby eliminating uninteresting beliefs.[40]

[39] In fact, it is hard to see how we could accept CA without thereby committing ourselves to CE, whatever our reason for accepting CA.

[40] This is not to say that virtuous forgetting necessarily implements CE perfectly: if the specific limitations on the subject's computational resources mean that she must forget more than is required by clutter elimination, she will necessarily eliminate some interesting beliefs; but if her memory system is

4 Is Human Forgetting Virtuous?

My thesis has two parts: first, that virtuous memory for finite cognizers will involve a certain pattern of forgetting; second, that forgetting in normal human memory approximates this pattern. The argument of Sects. 2 and 3 is meant to establish the first part of the thesis. The remaining question is to what extent the pattern of forgetting associated with human memory approximates that associated with virtuous memory. I will not be able to offer a decisive answer to this question here; but I can review work from psychology which suggests that the pattern of forgetting characteristic of normal human memory does indeed (imperfectly) approximate that associated with virtuous memory.

Both Bjork and his colleagues and Anderson and his colleagues have argued that the pattern of forgetting characteristic of normal human memory is adaptive.[41] There are important differences between the two approaches, but they share a common core idea: the memory system renders records inaccessible (in part) according to their retrieval history, effectively assuming that the history of use of a record predicts future need for the record; and the predictions of the memory system about the future needs for records are quite accurate. In other words: the pattern of forgetting mirrors the informational requirements imposed on memory by the subject and her environment.

Focussing on autobiographical memory, Bjork and Bjork develop a theory of disuse to account for the pattern of forgetting characteristic of normal human memory. I will not review the details of the theory here; it is sufficient for my purposes to describe the core of the approach. Summing up their approach, Bjork and Bjork write:

> In general, the theory of disuse ... says that the items in memory that are readily accessible to us are those items that we have been using (retrieving) lately. ... [T]hat will typically be adaptive. The items that have been retrieved frequently in the recent past will tend to be those items most relevant to our current interests, problems, goals, and station in life. On a statistical basis, those same items will be maximally relevant in the future as well. Items that have not been retrieved in the recent past, on the other hand, will tend to be those that are not as relevant to our current situation and, statistically, are not likely to be as relevant to our near future either. So, in general, those things that we are likely to need to recall in the near future will be accessible to us, and those things that are irrelevant or interfering or out-of-date will be inaccessible ... . (1988, pp. 285–286)

On this approach, our interests evolve according to a predictable pattern (presumably due in part to the predictable evolution of the informational

Footnote 40 continued
virtuous, she will preferentially forget clutter, thus minimizing the number of interesting beliefs that she forgets.

[41] Though the view that human cognition (including memory) is adaptive is widespread, it is not uncontroversial. See, e.g., the responses to Anderson 1991 in the same issue.

requirements of the environment):[42] given that we have frequently been interested in a record in the recent past, it is likely that we will continue to be interested in it in the near future. The predictability of the evolution of our interests means that the memory system can in principle anticipate our future interest in a given record: if we are interested in a record, we will retrieve it; retrieval history thus predicts future interest in a record. Thus if records are rendered inaccessible according to the frequency and recency with which they have been retrieved, the pattern of forgetting will correspond fairly closely to the informational requirements imposed on memory by the subject's interests: records in which the subject is no longer interested will tend to be rendered inaccessible, while records in which the subject continues to be interested will remain accessible. The theory of disuse proposes a particular mechanism by which the memory system can implement this pattern of forgetting, and claims that assuming that forgetting is regulated by this mechanism predicts the observed pattern of forgetting in human memory. In short, the theory suggests that we do in fact tend to forget approximately those records in which we are no longer interested.

Thus Bjork and Bjork's theory of disuse seems to suggest that the pattern of forgetting associated with normal human memory corresponds fairly closely to the pattern of forgetting associated (according to my argument) with virtuous memory. Something similar is suggested by the "rational analysis" of memory developed by Anderson and his colleagues (Anderson 1990). The rational analysis of memory suggests that the human memory system in effect takes its own specific computational limitations into account when determining what to forget. In Anderson's framework, retrieving a memory has a cost C (which reflects in part the time necessary for searching and considering the memory) (Anderson and Schooler 2000, pp. 557–558); if the memory is useful in the current context, retrieving it has a gain G. The problem facing the memory system is that of minimizing the costs of retrieval while maximizing the gains (Anderson and Schooler 2000, p. 558). The rational analysis of memory proposes that the memory system can do this because it can in effect "assign some probability P to a memory being relevant in advance of retrieving it" (Anderson and Schooler 2000, p. 558). According to Anderson, the human memory system is adaptive in the sense that it does in fact minimize the costs of retrieval while maximizing the gains:

> [A]n adaptive memory system would search memories in order of their expected utilities, $PG - C$, and stop considering memories when a probability P is retrieved such that $PG < C$. This predicts that people will be able to retrieve most rapidly memories that are most likely to be relevant to their current needs and not recall memories that are unlikely to be relevant. (Anderson and Schooler 2000, p. 558)

That is, the memory system is adaptive in the sense that it retrieves a given relevant record only if the cost of retrieving it does not exceed the probable gain of retrieving it. The trick is to figure out how to estimate P optimally, given the two sources of information available to the memory system, the history of use of the record and the

[42] See the discussion of the relativity of virtue to the normal human environment in Sect. 1 above.

cues provided by the subject's current context. The mathematics involved at this point get fairly complicated, but the basic idea is straightforward enough: for context, the system draws on associative strengths between cues and memories; for history, the system relies on frequency and recency of retrieval. Thus the rational analysis approach again seems to suggest that the pattern of forgetting characteristic of normal human memory approximates that associated with virtuous memory.

But in fact we cannot conclude that there is a very close correspondence between the two patterns. If forgetting is determined by retrieval history more or less in the manner suggested by theories of adaptive memory, then the memory system does indeed prefer to forget records in which the subject is no longer interested. But the sort of interest at issue in these theories can at best correspond to the first sort of interest invoked by my account, viz., subjective interest: retrieval, obviously, is triggered by queries that the subject actually sends to her memory, not by queries that she would send if she happened to think about certain questions or if she were more fully apprised of her interests; thus retrieval history can only predict future subjective interest, not future dispositional or objective interest.

The upshot is that, if the account of virtuous memory developed in Sect. 2 is right, normal human memory (adaptive memory) corresponds at best only loosely to virtuous memory: a virtuous memory system will preserve access to records in which the subject continues to be subjectively, dispositionally, or objectively interested; but an adaptive memory system will (unless the subject's subjective interests happen somehow to coincide with her dispositional and objective interests) forget many records which continue to be of dispositional or objective interest to the subject but which are not of subjective interest to her. We have two options at this point: either we insist that virtuous memory preserves access to records that continue to interest the subject in whatever sense (subjective, dispositional, or objective), in which case there will be only a fairly loose correspondence between normal human memory and virtuous memory, or we revise our account of virtuous memory so that it implies only that virtuous memory preserves access to records that continue to interest the subject in the subjective sense, in which case there will be a much tighter correspondence between normal human memory and adaptive memory.

The latter move might appear to be ad hoc, but it is on reflection not obvious whether we should allow dispositional and objective interests to play the role assigned to them in Sect. 2. There are two problems with the account of virtuous memory developed there. First: According to that account, a virtuous memory system will tend to retrieve records in which the subject continues to be interested in any of the three senses. Such a system will sometimes retrieve, in addition to records in which the subject is subjectively interested, records in which she has no subjective interest—records which do not answer any question that she actually cares about. Relying on the role of the memory system in the broader cognitive life of the subject to indicate the function of that system, it is at best unclear that retrieving such records is part of the function of memory—the subject will, after all, presumably make no use of a retrieved record in which she has, e.g., a merely dispositional interest. Second: Even if it would in some sense be desireable for the memory system to preserve access to records in which the subject has merely dispositional or objective interests, it is at best unobvious whether such a system is feasible. Merely dispositional or objective interests in general leave no internal traces (since the subject herself need in no way be aware of them), and thus it is hard to see how a memory system could be engineered to be sensitive to such interests. Thus I suggest that only subjective interests are after all relevant here: a virtuous memory system will prefer to forget subjectively uninteresting records.[43]

I have argued, against the default view, that a certain pattern of forgetting is associated with a virtuous memory system for any finite cognizer: given limited computational resources, forgetting is necessary to enable the system to achieve the balance of reliability, power, and speed appropriate for it given its function; given the necessity of sensitivity to interests for virtue, virtuous memory involves preferentially forgetting uninteresting records. Research on adaptive memory suggests that (depending on how we define the interests to which virtuous memory must be sensitive) the pattern of forgetting characteristic of normal human memory approximates that associated with virtuous memory fairly closely.

Acknowledgments Thanks to Louise Antony, Charles Clifton, Hilary Kornblith, Chris Lepock, Joëlle Proust, Jonathan Schaffer, John Sutton, two anonymous referees, and audiences at an APIC seminar at the Institut Jean-Nicod and the 2010 meeting of the Dutch-Flemish Society for Analytic Philosophy for comments. The preparation of this article was supported in part by the Agence Nationale de la Recherche, under the contract ANR–08.BLAN–0205–01.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale: Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14*, 471–517.

Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving & F. I. M. Craik (Eds.), *Handbook of memory* (pp. 557–570). Oxford: Oxford University Press.

Bjork, E. L., & Bjork, R. A. (1988). On the adaptive aspects of retrieval failure in autobiographical memory. In M. M. Grueneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 283–288). New York: Wiley.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 309–330). Hillsdale: Erlbaum.

Bjork, R. A., & Vanhuele, M. (1992). Retrieval inhibition and related adaptive peculiarities of human memory. In J. F. Sherry & B. Sternthal (Eds.), *Advances in consumer research* (Vol. 19, pp. 155–160). Provo: Association for Consumer Research.

Blustein, J. (2008). *The moral demands of memory*. Cambridge: Cambridge University Press.

Boyer, P. (2009). What are memories for? Functions of recall in cognition and culture. In P. Boyer & J. V. Wertsch (Eds.), *Memory in mind and culture* (pp. 3–28). Cambridge: Cambridge University Press.

Cherniak, C. (1986). *Minimal rationality*. Cambridge: MIT.

Dennett, D. C. (1987). *The intentional stance*. Cambridge: MIT.

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

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3*, 128–135.

[43] Alternatively, we might content ourselves with the loose correspondence between adaptive memory and virtuous memory that obtains if we insist that virtuous memory preserves access to records that continue to interest the subject in whatever sense, as long as the correspondence is not too loose: if it can be argued—I will not propose such an argument here— that normal humans are after all fairly good at bringing their subjective interests into line with their dispositional and objective interests, the correspondence will be better than I have been assuming that it is, though still imperfect.

Gigerenzer, G., Todd, P. M. and the ABC Research Group. 1999. *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Goldman, A. (1992). *Liaisons: Philosophy meets the cognitive and social sciences*. Cambridge: MIT.

Goldman, A. (1999). *Knowledge in a social world*. Oxford: Clarendon Press.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge: MIT.

Kraemer, P. J., & Golding, J. M. (1997). Adaptive forgetting in animals. *Psychonomic Bulletin and Review, 4*, 480–491.

Lackey, J. (2005). Memory as a generative epistemic source. *Philosophy and phenomenological research, 70*, 636–658.

Lepock, C. (2009). Unifying the intellectual virtues. *Philosophy and Phenomenological Research*. (Forthcoming).

Liao, S., & Sandberg, A. (2008). The normativity of memory modification. *Neuroethics, 1*, 85–99.

Luria, A. R. (1987). *The mind of a mnemonist*. Cambridge: Harvard University Press.

Marcus, G. (2008). *Kluge*. Boston: Houghton Mifflin.

McClelland, J., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex. *Psychological Review, 102*, 419–457.

McCloskey, M., & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–164). San Diego: Academic Press.

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences, 32*, 493–510.

Metcalfe, J. (1994). *Metacognition: Knowing about knowing*. Cambridge: MIT.

Michaelian, K. (2010a). Generative memory. *Philosophical psychology*. (Forthcoming).

Michaelian, K. (2010b). In defence of gullibility: The epistemology of testimony and the psychology of deception detection. *Synthese*. (Forthcoming).

Michaelian, K. (2010c). Is memory a natural kind? *Memory Studies*. (Forthcoming).

Millikan, R. G. (1984). *Language, though, and other biological categories*. Cambridge: MIT.

Parker, E. L., Cahill, L., & McGaugh, J. L. (2006). A case of unusual autobiographical remembering. *Neurocase, 12*, 35–49.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constaints imposed by learning and forgetting functions. *Psychological Review, 97*, 285–308.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. New York: Houghton Mifflin.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*, 610–628.

Sosa, E. (1991). *Knowledge in perspective*. Cambridge: Cambridge University Press.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*, 299–313.

Sutton, J. (1998). *Philosophy and memory traces: Descartes to connectionism*. Cambridge: Cambrige University Press.

Tulving, E. (1999). On the uniqueness of episodic memory. In L. G. Nilsson & H. J. Markowitsch (Eds.), *Cognitive neuroscience of memory* (pp. 11–42). Gottingen: Hogrefe and Huber.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381–391.

Wixted, J. T. (2007). Forgetting: It's not just the opposite of remembering. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 329–335). Oxford: Oxford University Press.

Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.