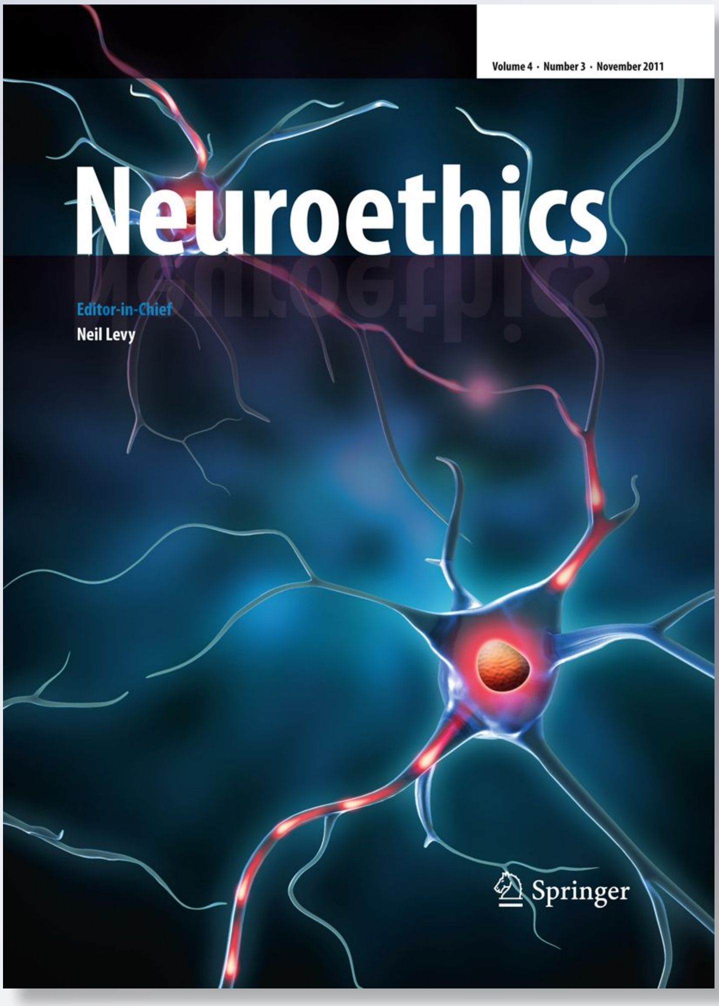
What is the "Cognitive" in Cognitive Neuroscience?

Carrie Figdor

Neuroethics

ISSN 1874-5490

Neuroethics
DOI 10.1007/s12152-012-9157-5



Neuroethics
DOI 10.1007/s12152-012-9157-5

ORIGINAL PAPER

What is the "Cognitive" in Cognitive Neuroscience?

Carrie Figdor

Received: 18 January 2012 / Accepted: 16 March 2012

Abstract This paper argues that the cognitive neuroscientific use of ordinary mental terms to report research results and draw implications can contribute to public confusion and misunderstanding regarding neuroscience results. This concern is raised at a time when cognitive neuroscientists are increasingly required by funding agencies to link their research to specific results of public benefit, and when neuroethicists have called for greater attention to public communication of neuroscience. The paper identifies an ethical dimension to the problem and presses for greater sensitivity and responsibility among neuroscientists regarding their use of such terms.

Keywords Cognitive neuroscience · Folk psychology · Research results · Translational implications · Public communication · Media ethics · Science reporting

Cognitive neuroscientists are increasingly required by funding agencies to link their research to specific potential outcomes of public benefit or interest [1, 2]. They are also increasingly aware of the need to communicate their research results more effectively to the public [3]. This paper argues that miscommunication and public misunderstanding of neuroscience results and implications stem to a significant degree from neuroscientists' failure to be sufficiently sensitive to the nature of the mental or cognitive concepts in terms of which they interpret their results and draw implications. In consequence, more effective communication of results and the drawing of justifiable translational implications depends in part on neuroscientists' willingness to assume greater responsibility for these choices. This problem can also affect collaborations with the social sciences and psychology and generates new neuroethical concerns.

C. Figdor (✉)
Department of Philosophy and Interdisciplinary Graduate Program in Neuroscience, University of Iowa,
260 English-Philosophy Building,
Iowa City, IA 52242, USA
e-mail: carrie-figdor@uiowa.edu

Implications of Using Mental Terms in Cognitive Neuroscience

The pressure to "sell" the broader impacts of one's research in order to get funded is not unique to cognitive neuroscience. Nor is cognitive neuroscience alone in contending with the difficulties of presenting complex research to the public. Principal dangers in both cases include the oversimplification of research results and public misunderstanding of the near- and long-term benefits of the research.

But cognitive neuroscience results have a unique character in terms of their potential impact on the public. Cognitive neuroscience is in a rare position within the sciences in that it is a bridging discipline between biology and the social and psychological sciences through its efforts to link the brain with the

Published online: 12 April 2012

mind. Because these findings have the potential to alter directly how human beings understand themselves, including their personal, moral and other social choices and relations, strong public interest in its findings is guaranteed. But these very implications also put cognitive neuroscientists in a unique position of responsibility regarding public misunderstanding, since they are directly aware of and have expert knowledge about the studies from which these results and implications are drawn. Science reporters are responsible for how they transmit cognitive neuroscience results and implications to the public, but cognitive neuroscientists are responsible for their choices of cognitive and mental terms to describe their results and implications to begin with [4]. Assuming their share of responsibility for avoiding miscommunication, I argue, involves greater sensitivity among cognitive neuroscientists to the potential for the cognitive or mental terms in which they routinely report their results to engender public misunderstanding and abet confusion.

In a recent article calling for improved public communication by neuroscientists, Illes et al. (op. cit.: 61) begin with the following remarks:

> Neuroscientists are faced with an important challenge. With the development of powerful new research tools, they are gaining a better understanding of the biology of the brain every day. At the same time, this progress is prompting many questions about the personal, social, moral and spiritual choices that humans make. These factors conspire to place increasing pressure on neuroscientists to discuss both their scientific research and the ethical implications of their findings.

What is not explicitly stated in this argument for improved neuroscience communication is the idea that the brain's operations are intimately connected to those of the mind as ordinarily conceived. Without this assumption there is no swift passage from more knowledge of the biology of the brain to insight into personal, social, moral and spiritual questions, typically posed in the familiar mental terms of folk psychology (e.g., self, love, guilt, and faith).

But in subsequent remarks the terms of this intimate relation shift to a link between neurology and *behavior*, such that it makes sense to describe the public as interested in "the neurological basis of individual and social behavior" (Illes et al: 61), rather than the neurological basis of ordinary mental phenomena. Of course, what is clear to Illes et al. and their scientific readership, but not necessarily to the public, is that the brain-mind link is cashed out in the laboratory by seeking associations between brain activity or deficits and observed behavior or deficits. (For brevity, reference to deficits will be assumed in what follows.) Much neuroscience research directly involves finding brain activity that can be reliably correlated with behaviors evoked in the performance of carefully designed tasks in controlled conditions. However, reports of these results and their implications in professional academic journals are routinely couched in cognitive or mental terms that refer to cognitive processes inferred directly or indirectly from the behaviors. Such cognitive inferences presumably justify reporting brain-behavior associations as brain-mind associations, as well as switching back and forth between the two.

Illes et al. can take all this for granted, since their intended scientific audience is similarly aware of the inferences that justify this use of cognitive or mental terms. But these uses of our everyday mental vocabulary or of closely aligned cognitive concepts must be handled with care, since they can unintentionally mislead the public. In what follows I identify three factors that can contribute to such miscommunication.

First, after over two decades of exploring the brain-mind link in normal humans using new non- or minimally-invasive imaging technologies (along with other methods), cognitive neuroscientists agree that this link is not going to be simple. To the contrary, research results reported in terms that link the brain directly with the mind as ordinarily conceived (e.g., [5, 6]) are frequently criticized within the scientific ranks, often for being new forms of phrenology ([7], Table 1 lists some of those criticized). But this criticism implicitly acknowledges that the use of mental terms to report results or characterize implications strongly suggests a close brain-mind connection that neuroscientists today (unlike the phrenologists) do not endorse. Saying that a structure or network is "involved in" a given mental capacity, or using the neologism "brain/mind", does little to eliminate this source of misunderstanding, and may in fact exacerbate it.

Second, personal, social, moral and spiritual concerns voiced in the vocabulary of folk psychology are formulated within a largely assumed context of some form of mental realism [8, 9], if not outright dualism. At the very least, the public is unlikely to think mental terms are convenient labels for patterns of behavior

[10] or that they don't really refer to anything [11]. However, the use of such terms to report cognitive neuroscience results and implications strongly suggests that realism about folk psychology has been justified by neuroscientific research. Of course, the vast majority of cognitive neuroscientists presumably do embrace realism about the cognitive or mental states to which they infer (despite the fact that the ubiquity of classical and operant conditioning methods in cognitive neuroscience experimental designs – from behavioral and imaging studies to connectionist modeling (e.g., [12]) – has made some raise the question of whether some form of behaviorism remains, or should remain, alive and well in neuroscience [13, 14].) But this global presumption of realism does not come with a specification of which, if any, mental states of folk psychology cognitive neuroscience will end up endorsing. For example, Lenartowicz et al. [15] propose a new methodology for mapping cognitive functions to brain structures using data-mining techniques over the task labels (e.g., "response selection") used in imaging studies. They conclude that the reliability (or not) of inferring to task labels from imaging results can show whether the cognitive functions thought to be involved in performing a task "are truly distinct componential entities or whether they emerge from the interactions of various systems in the brain and are therefore manifest only in the minds of cognitive scientists." A similar sense of "emergentism" is expressed by McClelland et al. [16], for whom the structured mental representations of classical computationalism – the cognitive architecture that reflects most closely the constructs of folk psychology – are "abstractions that are occasionally useful but often misleading: they have no real basis in the actual processes that give rise to linguistic and cognitive abilities or to the development of these abilities." Cognitive neuroscientists may *intend* to be realists about the mind as ordinarily conceived, but the routine use of ordinary mental terms to report results suggests that realism about folk psychology is already scientifically established.

However, the third, and most important, way in which the use of mental terms in cognitive neuroscience can contribute to miscommunication is the fact that cognitive neuroscientists often use these terms in conflicting ways that bear variable relationships to their ordinary meanings. The reasonable presumption that mental terms are being used in cognitive neuroscientific contexts with the same meanings they have in ordinary contexts unless explicitly stated otherwise is often violated. Cognitive inferences are made from an increasingly wide range of behaviors (verbal self-reports, repetitive bar-pressing, passive perception of photographs, questionnaire responses, directed saccades, startles, skin conductance responses, etc.) performed by human and non-human species (often rats and monkeys). These behaviors provide varied kinds of evidence, with variable degrees of justification, for various specific kinds of cognitive or mental states and capacities. In the absence of discipline-wide criteria for guiding cognitive inferences made from an ever-expanding basis of kinds of brain-behavior associations, it is left to individual researchers to determine the nature of the cognitive constructs inferred in a given study – that is, to choose which cognitive or mental concepts and terms are appropriate for interpreting and reporting their results and drawing implications. These choices often differ from study to related study in unannounced ways, even though the standard scientific practice of citing previous relevant work essentially depends on a presumption of shared meaning, not just shared forms of words, across studies.

It is true that the public can often be relied on to understand when ordinary terms are being used in non-standard, particularly metaphorical, ways in science: for example, everyone is aware that the subtitle of the book "Suprachiasmatic Nucleus: the mind's clock" [17] includes a metaphor. But when a cognitive neuroscientist uses ordinary mental terms to report her research, the public – including science reporters who read peer-reviewed articles and attend the discipline's conferences – cannot be expected to know or even suspect that what one cognitive neuroscientist means may not be what another means, and that what both mean may depart from ordinary meanings in subtle but important ways. For example, a study on the neural correlates of love by Fisher et al. [18] – part of the first citation tree provided below – appears in a special issue of the *Journal of Comparative Neurology* entitled "The anatomy of the soul". At the very least, the term "soul", unlike "clock", is not clearly metaphorical or non-literal when used in a science that is trying to link the brain to the mind as ordinarily conceived.

Perhaps the best evidence of this varability in usage comes from critically examining uses of the same or related cognitive terms across peer-reviewed studies via citation trees, diachronically and synchronically. The papers discussed below are linked either directly (paper A cites paper B) or else indirectly with one degree of separation (B in turn cites C); papers not in the trees are noted as such. (Quotation marks flag target terms and are neither scare-quotes nor (by themselves) use-mention quotes; use-mention quotes are preceded by the phrase "the term"; no quotation marks indicates a use of a target term to refer to an operation of the mind as ordinarily conceived.) The first tree, ending with [19], involves the terms "reward" and "motivation"; the second, ending with [20], involves "fear", "memory" and "fear memory". Within each tree, ordinary mental terms are used to refer to operations whose relations to the mind as ordinarily conceived vary – the relation may be identity, but need not be – and, more importantly, are not clearly defined or their departure from ordinary meanings not stressed. An indicator of fluctuation is the introduction of new mental terms in a study that are semantically related to the target concepts only if their meanings are folk-psychological, even when behavioristic or otherwise more restricted definitions are given or adopted from prior cited papers that do not include or emphasize the richer vocabulary.

First case Olds and Milner [21] reported the discovery of "reward" centers in the septal area of the central forebrain and other structures as revealed in operant conditioning protocols involving rats pressing bars to self-administer mild electrical charges via implanted electrodes. They report their results to peers in orthodox behaviorist fashion, where a "reward" (or positive reinforcer) is defined as a stimulus associated with increased frequency of response, "stimulation" is an electrical charge, and stimulation in these brain areas is "rewarding in the sense that the experimental animal will stimulate itself in these places frequently and regularly for long periods of time if permitted to do so." Olds [22] reports these results to the public in terms connoted by their everyday meanings: the same regions are called "pleasure centers", "pleasure" is described a feeling or experience (and is ascribed to a non-human animal), electrical stimuli are called "rewarding", and "the [rat's] pleasure of stimulating itself electrically" is ranked as "more satisfying than the usual rewards of food, etc." Wise [23] describes Olds (op.cit.) as using the term "pleasure" heuristically, which seems correct given that Olds (op.cit.) also credits B.F. Skinner with refining the methods for measuring pleasurable and painful feelings. But this heuristic use is predicated on a meaning relation to reward as ordinarily conceived, not as behaviorally defined.

Later, Kawagoe et al. [24] found that "memory-related" neural responses were modulated by "expectation of reward, either as enhancement or reduction of response." Their study combined operant conditioning of monkeys with single-cell recording of caudate nucleus neurons, but they discuss "reward" – a liquid given to the monkeys that increased the frequency of correct saccades to a target – in an expanded conceptual context involving concepts of "expectation" and "memory" that are not behaviorally defined. A "memory-related" neural response is clearly defined as sustained phasic activity in recorded neurons that started at least 200 milliseconds after cue onset and ended before or with a saccade, but the term "memory" is not defined.

By the time of Aron et al. [19], human subjects who self-report being madly in love are imaged while passively viewing photographs of their loved ones (see also Bartels and Zeki [25]). Subjects did not perform a task the correct performance of which yielded more opportunities to view the photographs; thus, they could not display "motivation" to attain this "reward" as these constructs were previously explicitly defined in the citation tree. Nevertheless, the researchers report finding that subcortical systems associated with "motivation" to acquire a "reward" in prior studies are among those associated with romantic love. The septum is described as a region "found to be rewarding during electrical self-stimulation [21]", and Kawagoe et al. (op.cit.) is one of several papers cited as showing that "the caudate nucleus plays a major role in reward and motivation in the mammalian brain …"

In short, a neural system initially associated with "reward" in the sense of stimuli correlated with increased frequency of a measured response is later associated – whether instead or in addition is not clear – with feelings of pleasure inferred by connotation from the ordinary meaning of the term "reward" and with responses inferred from passive viewing of visual stimuli classified as "positive" due to a cognitively-mediated relation between recognized photographic content and a human subject's self-report of a pleasurable complex folk-psychological emotion. As a result, even if the septum and caudate nucleus are indeed part of a "reward and motivation" system, it is unclear what is meant by calling this a "reward and motivation" system or, in nonlinguistic terms, what relations the operations inferred in a given study have to reward and motivation as ordinarily conceived. Behavioral definitions of "reward" and "motivation" appear increasingly vestigial but are not clearly explicitly replaced or augmented.

Second case Blanchard and Blanchard [26] distinguished crouching as a rat's response to "fear" because rats conditioned with shocks crouched when later returned to the shock chamber without being shocked and did not crouch once removed from it. Misanin et al. [27] used lick rates (the time required for the rat to make its first 100 licks of water in the experimental apparatus) to measure "fear", and conditioned rats to associate a white noise with a footshock. Rats given electroconvulsive shock 24 hours after "fear"-conditioning but immediately after a presentation of white noise exhibited lick rates at pre-conditioning levels. They inferred that the rats exhibited a "memory loss of a fear response" or "retrograde amnesia", defined as "impaired retention of responses learned shortly before electroconvulsive shock (ECS) stimulation".

It is important to note the cognitive-inferential difference between these two very similar conditioning studies. In one, exhibiting a conditioned response to aversive stimuli after a delay justifies an inference to "fear", while in the other not exhibiting this kind of response justifies an inference to a "memory loss". The latter inference introduces a mental term that in ordinary contexts refers to a kind of mental state with representational content – the content ascribed or self-ascribed when (e.g.) an accident victim reports remembering that a silver SUV crashed into the car. ("Fear" also ordinarily has as its object something perceived as aversive – e.g., footshock and, after conditioning, white noise.) Fully robust declarative memory, complete with self-reference (e.g., "I remember that this noise was followed by a footshock") is presumably not intended, but it follows that the term "memory" does not have its primary folk-psychological meaning. What it does mean is not clear. Reference to a "fear memory" may just be a roundabout way of saying that the animal was conditioned. A more robust interpretation (but one that still falls short of involving self-reference) is that it is a "memory" of an association, or of a series of episodes, that includes (or include) an aversive relatum – roughly, either "This noise was followed by footshock" or "This noise was accompanied by feeling fear". Either way, a "fear memory" so understood ordinarily requires further cognitive processing to activate a somatic response, and so is more than mere "fear" conditioning.

This ambiguity in "fear memory" extends to what is inferred from extinction, since the latter cognitive inference depends directly on what is inferred from acquisition. Extinction is when the conditioned stimulus (CS) is presented without the unconditioned stimulus (US) until the conditioned response (CR) is no longer observed. After extinction Blanchard and Blanchard's rats no longer feel "fear", but what Misanin et al.'s rats undergo when they cease to exhibit the conditioned response is unclear (although they call it "memory loss of a fear response"). They may also just lose the conditioning. More robustly interpreted, they may lose the "memory" of an association or a sequence (roughly specified above), or gain a new "memory" or other kind of new learning (itself unspecified), or cease engaging in the additional processing needed to yield a somatic response from the acquired "fear memory". However, the term "retrograde amnesia" can refer to what is inferred from the rats' ceasing to exhibit the conditioned response only if the cognitive term used to label the initial result of conditioning is implicitly assumed to have a meaning close to its ordinary meaning.

Once the term "fear memory" is introduced, subsequent related brain-behavior and conditioning studies freely use these terms and introduce others associated with them in ordinary contexts (e.g., "memory retrieval", "anxiety", "traumatic memory") to report results and implications. In Nader et al. [28], the lateral and basal amygdalae of "fear"-conditioned rats were infused with anisomycin (a protein synthesis inhibitor) immediately after CS presentation or without CS presentation. "Reactivation of fear memory" was inferred from the infused rats' immobility in the CS-presentation condition, while infused rats in the other condition showed "amnesia" inferred from the loss of this conditioned response. Phelps et al. [29] performed an fMRI study to see if known correlations between ventromedial prefrontal cortex (vmPFC) and the amygdala and "fear" learning and extinction in non-human animals could be replicated in humans. Human subjects were presented with blue squares as the CS+(accompanied by mild wrist-shock in conditioning trials) and yellow squares as the CS- (never accompanied by wrist-shock). Acquisition and extinction were measured by differential skin conductance response (SCR). The fMRI scans showed correlations with acquisition and extinction (e.g., the blood-oxygen-level-dependent (BOLD) signal increasing in acquisition and decreasing in extinction) in the amygdala and, within medial PFC, the subgenual anterior cingulate. However, since the nature of the cognitive operations inferred in acquisition and extinction are unspecified, it is hard to know what relation these operations picked out by the terms "fear" and "memory" have to fear and memory as ordinarily understood.

Clues from other papers about the nature of what is being (or may be) inferred conflict. Milad et al. [30] specify that the "memory" in conditioning is "a CS/US association" and that in extinction as "a Cs/no-US association", although the content of a memory of a noise that is not paired with anything cannot be a memory of an association. Looking outside this citation tree, if "an ability to link distinct experiences can be a basis for episodic memory" [31], then a "fear memory" of an association may be an episodic memory of a white-noise event followed by a footshock event (or a fear-feeling event) or an operation involved in episodic memory, where the critical link to memory as ordinarily conceived is labeled by the phrase "involved in". Alternatively (and also outside this citation tree), if we distinguish recognition memory ("judging an object as familiar") from recollecting a learning episode, and acknowledge that this distinction "is difficult if not impossible to apply to experimental animals", the "fear memory" might only be a recognition memory, and not a memory of an association or associative memory at all [32].

In addition, over time the concept "memory" has been generalized to the point where it is now defined as "a stored representation of experience" or, when referring to the capacity, "the ability of living organisms to retain and utilize acquired information" ([33], also not in this citation tree). If so, to say the hippocampus and vmPFC are involved in "memory" is to say they are involved in storing and using information acquired from experience, which is true of the rest of the brain as well. This weakened definition has a critical, if unintentional, consequence with respect to public understanding of the cognitive neuroscience of "memory". It enables cognitive neuroscientists to report their results and implications using the same term across studies whose behavioral bases for making cognitive inferences to "memory" can vary significantly, and hence whose relation to memory as ordinarily understood can also vary. This variance may help explain why researchers now agree that "memory" (at least in neuroscience contexts) "is not a single entity or concept" (Schacter op. cit.). (In the same vein, [34] distinguishes "autonomic arousal as classically conceived" from "a more information-processing concept of arousal that construes it as interruption of ongoing processing to gather more information (more aligned with orienting).")

In the meantime, research into "fear memory" has continued apace. Alvarez et al. [35] sought to replicate in humans findings with rodents that showed that extinction erases "the memory of the fear response" or "the memory of the fear learning" – specified as being of the CS-US association (roughly, "This tone was followed by a shock", setting aside an interpretation that involves self-reference). They conditioned human subjects by pairing tones and individually-determined moderately painful wrist-shocks, and used bursts of white noise as a startle stimulus to elicit "fear-potentiated startle", defined (in a standard way) as "the increase in startle reactivity when elicited in the presence of a CS previously paired with an aversive US". They found that in extinction subjects used contextual cues to acquire a distinct new mental state that has the same CS as part of its content but without the "fear". Roughly, the acquired "fear memory" was "This tone in this context was followed by wrist-shock" and the mental state acquired in extinction was "Here's the same tone in this other, safe, context". In this case, the nature of "fear", rather than "memory", is unclear. The startle reflex is described both as a measure of "fear" and as "an established measure of fear and anxiety", "contextual anxiety" is operationally defined as the magnitude of startle during intertrial interval (ITI) as measured by electromyographic (EMG) eyeblink and SCR, and subjects are asked to report "overall levels of anxiety" after renewal tests – even though anxiety-potentiated startle, unlike "fear"-potentiated startle, is not a widely referenced construct.

Finally, Schiller et al.'s [20] study of "reconsolidation of fear memory" in human subjects sought evidence that "updating a fear memory with non-fearful information, provided through extinction training, would rewrite the original fear response and prevent the return of fear." In this study, a fully robust notion of "memory" is required: their results and implications depend essentially on the precise content of the subjects' "fear memories", and the contents ascribed appear to depend on our ordinary folk-psychological methods for ascribing mental states to others. Similar to Phelps et al. (op.cit.), views of differently colored squares were paired with uncomfortable but not painful wrist-shocks, and "fear" was measured by skin conductance response. Heightened SCR following CS presentation justified an inference to "reactivation of fear memory" or being "reminded of the fear memory". Subjects showed "reinstated fear responses" to the CS that was not "reminded" during extinction (a red or a green square), but no "recovery of fear" when presented with the "reminded" CS (a yellow or an orange square). In other words, the content of the "renewed fear memories" is assumed to be of an association between a square of a specific color and wrist-shock. When this "fear memory" is "reminded" during extinction, it may be that the "memory" is lost, or that a new "memory" or other new mental state with different content is formed, or that the additional processing that leads from the original "fear memory" to a somatic response is not made. Whichever cognitive inference is made, the terms "fear" and "memory" must refer to operations of the mind as ordinarily conceived in order for their results to imply that using the reconsolidation window to "rewrite emotional memories" could have application in treating anxiety disorders.

Some researchers may think Schiller et al. are unjustified in inferring to fear – e.g., that mild wrist-shocks are merely annoying, or that SCR is not a good measure of fear. This criticism is inappropriate if Schiller et al. really do mean by "fear" something like what Blanchard and Blanchard or Misanin et al. mean. But in that case the study's relevance to clinical anxiety is attenuated. The appropriate criticism is that the nature of "fear" in their study and its relation to fear as ordinarily understood, with its close connection to anxiety and clinical implications, has not been specified.

To be clear, I do not claim that the cognitive inferences in individual studies are unjustified, that the studies are flawed methodologically, that vmPFC and the amygdala are not involved in emotion processing or extinction, that humans and other species do not share cognitive constructs, or that behavioral and physiological measures of complex mental phenomena are unjustified. Even if there are methodological reasons to criticize some studies – e.g., using SCR to measure "fear" even though SCR is widely regarded as a "nonspecific measure of arousal that reflects orienting to a stimulus as a function of relevance and not necessarily its emotional significance" (Alvarez et al. op.cit.; Adolphs op.cit.) – methodological unsoundness in a study is not cognitive-inferential variability and ambiguity of meaning across studies. The issue here is the consequences for public understanding of neuroscience results and implications when, first, there is a lack of systematic relationship between the ordinary meanings of mental terms regularly used to report these results and draw these implications and the nature of the cognitive operations picked out by these terms in neuroscience contexts, and, second, when these semantic gaps are not clearly stressed. Anyone not deeply familiar with the experimental protocols and habitual linguistic usage within neuroscience is unlikely to realize that these variations in meaning exist, given that peer-reviewed papers, not just the popular science press, couch their results and implications in the same ordinary mental terms. This usage, of course, is part of what makes cognitive neuroscience research so compelling to the public in the first place.

Collaborations with social scientists and psychologists in research studies with a cognitive neuroscientific component (e.g., [36–39]) may also be affected by this problem. Social scientists and many psychologists make generous use of folk-psychological terms: preferences, decisions, cognitive biases, theory of mind, false memories and so on play essential, non-heuristic explanatory and predictive roles in their theories. In their own research, some of these scientists may face a similar issue of employing ordinary mental terms to refer to constructs inferred in laboratory contexts without clarifying that these terms may not be being used as they would in non-laboratory settings. In collaborations, the potential for confusion is compounded if collaborators disagree about the nature of the processes being inferred. To the extent that such disagreement exists but is not explicit, uses of mental terms to report results of a collaborative study will be ambiguous.

Implications for Neuroethics

Those who have delineated the concerns of neuroethics (e.g., [40, 41]; Farah op.cit., [42, 43, 44]) have similar rosters of issues that generally fall into three categories. (1) What are the implications of new knowledge about the brain (or CNS) for the moral status and ethical treatment of biological creatures, in particular those with cognitive capacities? Issues here involve personal responsibility, access to and uses of brain information, the scope of permissible brain modification, and animal rights. (2) How may advances in neuroscience affect non-scientific conceptions of mentality, morality, and humanity (or personhood in general)? Issues here involve questions about free will, the nature of the self, the conceptual analysis of moral judgment, decision making and knowledge, and the scientific viability of folk-psychological categories and the cognitive models based on them. (3) What research practices may be ethically pursued in neuroscience? These issues involve extending standard questions in bioethics regarding the ethical pursuit of biological research to the special case of neuroscience.

If my analysis is correct, the category (2) issues – especially but not limited to the scientific fate of folk psychology – are matched by a new, associated, category (3) issue: What are the ethical ways in which cognitive neuroscientists can interpret and report their research results and draw their translational implications? The gap between the ordinary meanings of mental terms and what these terms may mean in cognitive neuroscientific contexts raises the issue of foreseeable potential public misunderstanding. The ethical dimension of this issue can be seen by considering the following question: Why *should* cognitive neuroscientists highlight the differences between what is justifiably inferred in the laboratory and what the public may think has been inferred if it takes at face value the familiar mental vocabulary in which cognitive neuroscience research results are routinely couched? New requirements for specifying translational impacts provide pragmatic reasons to prioritize being more sensitive to the language used to report cognitive neuroscience results and implications. But whether neuroscientists *should* give this issue equal or more weight than other professional goals is a matter of basic ethical judgments regarding how the public should be treated. Moreover, individual scientists may weigh these values differently and act accordingly unless discipline-wide guidelines for the uses of mental terms or concepts are adopted.

Conclusion

It is an open question exactly what kind of bridge cognitive neuroscientists are building between the brain and the mind as ordinarily conceived. In the meantime, the routine use by cognitive neuroscientists of ordinary mental terms to describe the results and implications of their research is a source of potential public confusion and misunderstanding. As a result, cognitive neuroscientists do not get to use mental concepts for free. Since their results and implications are of deep personal, social, moral and spiritual import, the use of ordinary mental terms to report them requires sensitivity to how such terms are likely to be understood. Choosing to use such terms is a morally relevant choice.

Some neuroscientists may be taking steps to raise awareness of aspects of the problem among neuroscientists [45]. But while internal discussion goes on, each neuroscientist might consider the following smell test when she writes up her results or translational implications using terms from our everyday mental vocabulary: would a randomly chosen person infer to this kind of cognitive operation from the behavior observed in the study? If the answer is not clearly yes, it would seem that the minimal ethical thing for her to do would be to make the difference in meaning explicit in her research papers and translational implication statements, and to emphasize this difference when talking to science reporters.

It is possible that the more open cognitive neuroscientists are about divergence in meaning, the less the public may regard their results as illuminating personal, social, moral and spiritual questions. For example, it is one thing to use animal models for investigating neurodegenerative diseases that affect cognition and another to pick out what is inferred from their behavior using folk-psychological terms. On the other hand, because of its compelling public profile, by facing the problem cognitive neuroscientists may spearhead efforts to improve general education about the nature of science. In journalism, the ethical solution to news reporting in the face of ignorance is transparency about one's sources and careful writing that openly reflects the news report's degree of reliability [46]. Cognitive neuroscientists may consider adopting a similar ethical code regarding the way they describe, and communicate to the public, the results and implications of their research into the still unsettled brain-mind link. Even if more openness leads to an initial loss of public interest or even trust, in the longer term cognitive neuroscience will benefit from being more clear about what it is doing.

Acknowledgements My deepest thanks go to Bradley Taber-Thomas for extensive input on early drafts of this paper and Colin Klein for astute comments on the final draft. Thanks also to Daniel Tranel, Tom Brashers-Krug, and Justin Feinstein for comments and references, and to audiences at presentations of drafts at the University of Iowa Interdisciplinary Graduate Program in Neuroscience, the Iowa Philosophical Society 2010

annual meeting, and the Southern Society for Philosophy and Psychology 2011 annual meeting.

References

1. NIH Division of Program Coordination, Planning and Strategic Initiatives. Translational Research. http://commonfund.nih.gov/clinicalresearch/overview-translational.aspx. Accessed 28 March, 2012.
2. Maienschein, J., M. Sunderland, R.A. Ankeny, and J.S. Robert. 2008. The ethos and ethics of translational research. *The American Journal of Bioethics 8: 43–51.*
3. Illes, J., M.A. Moser, J.B. McCormick, E. Racine, S. Blakeslee, A. Caplan, E.C. Hayden, J. Ingram, T. Lohwater, P. McKnight, C. Nicholson, A. Phillips, K.D. Sauvé, E. Snell, and S. Weiss. 2010. Neurotalk: improving the communication of neuroscience research. *Nature Reviews Neuroscience* 11: 61–69.
4. Hamilton, Jon. 2011. Computers one step closer to reading your mind. *National public radio* (All Things Considered; Robert Siegel, host), March 11 broadcast
5. Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.
6. Quian Quiroga, R., L. Reddy, G. Kreiman, C. Koch, and I. Fried. 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435: 1102–1107.
7. Poldrack, R.A. 2010. Mapping mental function to brain structure: how can cognitive neurimaging succeed? *Perspectives on Psychological Science* 5(6): 753–761.
8. Fodor, J. 1987. *Psychosemantics*. Cambridge: MIT Press.
9. Griffiths, T.L., N. Chater, C. Kemp, A. Perfors, and J.B. Tenenbaum. 2010. Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences* 14: 357–364.
10. Dennett, D. 1987. *The intentional stance*. Cambridge: MIT Press.
11. Churchland, P. 1981. Eliminative materialism and the propositional attitudes. *The Journal of Philosophy* 78(2): 67–90.
12. Gluck, M.A., M. Meeter, and C.E. Myers. 2003. Computational models of the hippocampal region: linking incremental learning and episodic memory. *Trends in Cognitive Sciences* 7 (6): 269–276.
13. Machamer, Peter. 2009. Learning, neuroscience, and the return of behaviorism. In *The Oxford Handbook of Philosophy and Neuroscience*, ed. Bickle John, 166–176. New York: Oxford University Press.
14. Uttal, W. 2001. *The new phrenology*. Cambridge: MIT Press.
15. Lenartowicz, A., D.J. Kalar, E. Congdon, and R.A. Poldrack. 2010. Towards an ontology of cognitive control. *Topics in Cognitive Science* 2: 678–692.
16. McClelland, J.L., M.M. Botvinick, D.C. Noelle, D.C. Plaut, T.T. Rogers, M.S. Seidenberg, and L.B. Smith. 2010. Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences* 14: 348–356.
17. Klein, D.C., R.Y. Moore, and S.M. Reppert (eds.). 1991. *Suprachiasmatic nucleus: the mind's clock*. New York: Oxford University Press.
18. Fisher, H., A. Aron, and L.L. Brown. 2005. Romantic love: an fMRI study of a neural mechanism for mate choice. *The Journal of Comparative Neurology* 493: 58–62.
19. Aron, A., H. Fisher, D. Mashek, G. Strong, H. Li, and L.L. Brown. 2005. Reward, motivation and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology* 94: 327–337.
20. Schiller, D., Marie-H Montfils, C.M. Raio, D.C. Johnson, J.E. LeDoux, and E.A. Phelps. 2010. Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* 463: 49–53.
21. Olds, J., and P. Milner. 1954. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology* 47(5): 419–427.
22. Olds, J. 1956. Pleasure centers in the brain. *Scientific American* 195(4): 105–116.
23. Wise, R.A. 1996. Neurobiology of addiction. *Current Opinion in Neurobiology* 6: 243–251.
24. Kawagoe, R., Y. Takikawa, and O. Hikosaka. 1998. Expectation of reward modulates cognitive signals in the basal ganglia. *Nature Neuroscience* 1(5): 411–416.
25. Bartels, A., and Zeki, S. 2000. The neural basis of romantic love. *NeuroReport 11*(17): 3829–3834.
26. Blanchard, R.J., and D. Caroline Blanchard. 1969. Crouching as an index of fear. *Journal of Comparative and Physiological Psychology* 67(3): 370–375.
27. Misanin, J.R., R.R. Miller, and D.J. Lewis. 1968. Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science* 160 (3827): 554–555.
28. Nader, K., G.E. Schafe, and J.E. LeDoux. 2000. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* 406: 722–726.
29. Phelps, E.A., M.R. Delgado, K.I. Nearing, and J.E. LeDoux. 2004. Extinction learning in humans: role of the Amygdala and vmPFC. *Neuron* 43: 897–905.
30. Milad, M.R., C.I. Wright, S.P. Orr, R.K. Pitman, G.J. Quirk, and S.L. Rauch. 2007. Recall of Fear Extinction in Humans Activates the Ventromedial Prefrontal Cortex and Hippocampus in Concert. *Biological Psychiatry* 62(62): 446–454.
31. Eichenbaum, Howard. 2004. An information processing framework for memory representation by the hippocampus. In Gazzaniga, Michael S. 2004 (ed). *The cognitive neurosciences III*. Cambridge: MIT Press/CogNet Library
32. Squire, Larry R., Robert E. Clark and Peter J. Bayley. 2004. Medial temporal lobe function and memory. In Gazzaniga, Michael S. 2004 (ed). *The Cognitive Neurosciences III*. Cambridge: MIT Press/CogNet Library
33. Schacter, Daniel S. (2004). Memory: introduction. In Gazzaniga, Michael S. 2004 (ed). *The Cognitive Neurosciences III*. Cambridge: MIT Press/CogNet Librar
34. Adolphs, R. 2010. What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences* 1191(1): 42–61.
35. Alvarez, R.P., L. Johnson, and C. Grillon. 2007. Contextual-specificity of short-delay extinction in humans: renewal of fear-potentiated startle in a virtual environment. *Learning and Memory* 14: 247–253.

36. Fehr, E., and C.F. Camerer. 2007. Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences* 11: 419–427.
37. McClure, S.M., J. Li, D. Tomlin, K.S. Cypert, L.M. Montague, and P. Read Montague. 2004. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44: 379–387.
38. Harris, L.T., and S.T. Fiske. 2006. Dehumanizing the lowest of the low: neuro-imaging responses to extreme outgroups. *Psychological Science* 17: 847–853.
39. Ochsner, K.N., and M.D. Lieberman. 2001. The emergence of social cognitive neuroscience. *American Psychologist* 56 (9): 717–734.
40. Roskies, A. 2002. Neuroethics for the new millenium. *Neuron* 35: 21–23.
41. Roskies, A. 2009. What's "Neu" in Neuroethics? In *The Oxford Handbook of Philosophy and Neuroscience*, ed. J. Bickle, 454–470. New York: Oxford University Press.
42. Farah, M.J. 2005. Neuroethics: the practical and the philosophical. *Trends in Cognitive Sciences* 9(1): 34–40.
43. Illes, J., and S. Bird. 2006. Neuroethics: a modern context for ethics in neuroscience. *Trends in Neurosciences* 29(9): 511–517.
44. Farah, M.J. 2002. Emerging ethical issues in neuroscience. *Nature Neuroscience* 5(11): 1123–1129.
45. Berridge, K.C., and T.E. Robinson. 2003. Parsing reward. *Trends in Neurosciences* 26(9): 507–513.
46. Kovach, B., and T. Rosenstiel. 2001. *The elements of journalism*. New York: Three Rivers Press.