

argue that our ordinary attributions of intentions are not only frequently false, but also explanatorily otiose (Churchland 1988).

Some philosophical defenders of folk psychology dismiss such arguments as irrelevant to the probity of the belief-desire framework: Explanation within such a framework, they say, is different from causal explanation, and can be justified independently of any causal account of the production of action (Blackburn 1986; McDowell 2004). I am, however, staunchly naturalistic in my approach to the mind, and so am committed to the continuity of philosophical and empirical work. I therefore recognize the in-principle relevance of the work in question and the seriousness of the challenge it poses. I am very glad, therefore, for the critique N&S offer. I would like to highlight some aspects of their critique that I find particularly germane to the philosophical debate, and then complement the critical points made by the authors with considerations that raise additional questions about the work in question.

N&S point out that Nisbett and Wilson's landmark "stocking study" (Nisbett & Wilson 1977), alluded to earlier, fails to meet at least one of their criteria of adequacy for assessments of awareness, namely, relevance. Although spatial position was correlated with subjects' choices, Nisbett and Wilson illegitimately presume that spatial position per se was causally relevant to their subjects' choices. But this inference neglects the possibility that subjects were running a left-to-right sequential evaluation of the stockings and operating with the rule "if the next one is as good as the previous one, go with the next one." Indeed, Nisbett and Wilson themselves report evidence that suggests that subjects were doing something like this. I find this point especially significant for philosophy, because it underlines the importance of taking seriously cognitive states and processes as independent variables in the production of behavior, variables that must be studied and controlled for. This is as against the strict behaviorist model (which seems to have more vitality in philosophy than in psychology), which only considers publicly available factors—observable stimuli and behavioral responses—and the reductionist/eliminativist model, which says that it is otiose to posit states at any level of abstraction above the neurophysiological level.

But it's one thing to say that the proximal causes of the subjects' choices were *cognitive* and another to say that they were *introspectible*. If N&S are correct about the cognitive procedure the subjects were utilizing, why did the subjects not report that? Why did they insist that their choices were based on the superior quality of the stocking they chose? I have a hypothesis: The set-up of this experiment is a virtual invitation to confabulation. Since there is no good basis for preferring any one sample to any other, subjects will, typically, not be able to cite any such basis. Hence *any* reason proffered by the subject is going to be wrong. But what does a subject's behavior in this sort of circumstance tell us about the accuracy of introspection in cases in which the subject *does* have a reason for acting as he or she does?

What I am suggesting is that Nisbett and Wilson were investigating introspective awareness under *degraded conditions*. In general, it cannot be assumed that the way we solve problems in normal conditions is the same as the way we solve them in degraded conditions. (Consider the very different visual processes activated in daylight and in low light.) Inferences about the unreliability of a certain cognitive process in degraded conditions should not be taken as evidence that the same process is unreliable in normal circumstances. (If we assessed color vision by looking at its operation in low light, we'd conclude that we are terrible at judging colors.) It could well be, therefore, that introspection is highly reliable when our choices and actions are the result of reasons—that is, when there are reasons there to be introspected—but that we have to employ other methods of explaining our own behavior—perhaps, as Nisbett and Wilson suggest, theoretical inference—in cases where introspection finds nothing there. Of course, it would be very difficult to design an experiment to test the accuracy of introspection in what I'm assuming are the

circumstances optimal for its operation. We would have to have circumstances in which the agent *has* a reason, and *we* know what it is. And it's hard to see how those conditions could be operationalized; it's much easier to set things up so that the agent *has* to be wrong. But of course, scientists should not be looking under the corner lamppost for watches dropped in the middle of the street.

I called the stocking comparison set-up a case of "degraded conditions." The degradation here is the absence of any reason in the agent's mind for introspection to detect. Other kinds of sub-optimality include *hard cases*—cases where there are or might be rational bases for decision, but these bases do not readily determine the best course—and *marginal cases*—cases where there are non-rational factors, such as emotional responses, that feed into the agent's decision. Asking for an agent's reasons in any of these circumstances is likely to provoke a state of mind similar to those that are called cases of *dumbfounding* in the literature on the psychology of moral judgment—cases in which subjects report strong moral judgments for which they offer no compelling moral justification. Accordingly, I would make a similar criticism of work that attempts to draw inferences about our ordinary moral reasoning from the responses subjects make in such cases: It is methodologically unsound to draw conclusions about our ordinary moral decision making from post hoc rationalizations of judgments about hard or marginal cases.

Maybe it helps to be conscious, after all

doi:10.1017/S0140525X13000630

Roy F. Baumeister,^a Kathleen D. Vohs,^b and E. J. Masicampo^c

^aDepartment of Psychology, Florida State University, Tallahassee, FL 32306-4301; ^bCarlson School of Management, University of Minnesota, Minneapolis, MN 55455; ^cDepartment of Psychology, Wake Forest University, Winston-Salem, NC 27109.

baumeister@psy.fsu.edu vohsx005@umn.edu

masicaej@wfu.edu

<http://www.psy.fsu.edu/faculty/baumeister.dp.html>

<http://www.carlsonschool.umn.edu/faculty-research/vohsx005/>

[Kathleen_D_Vohs.aspx](http://www.wfu.edu/~masicaej/Kathleen_D_Vohs.aspx)

<http://www.wfu.edu/~masicaej/>

Abstract: Psychologists debate whether consciousness or unconsciousness is most central to human behavior. Our goal, instead, is to figure out how they work together. Conscious processes are partly produced by unconscious processes, and much information processing occurs outside of awareness. Yet, consciousness has advantages that the unconscious does not. We discuss how consciousness causes behavior, drawing conclusions from large-scale literature reviews.

Science proceeds by approaching the truth gradually through successive approximations. A generation ago, psychologists began to realize that the conscious control of action had been overestimated, and many began to search for unconscious processes. The pendulum has now swung the other way, as Newell & Shanks (N&S) have shown in their valuable corrective to the excesses of that view. Now it is the unconscious effects that have been overestimated.

We think the way forward for psychological theory is to stop pitting conscious against unconscious and instead figure out how the two work together. Plainly, there is plenty of processing of information that occurs outside of awareness. Likewise as plainly, consciousness has advantages and can accomplish things that the unconscious cannot.

N&S propose their lens model with multiple stages of processing. They rightly criticize the tendency to claim that some outcome is unconscious based on showing that only one of the five steps is unconscious. Still, the fact that some steps are

unconscious is a genuine contribution that will need to be included in the eventual, correct account of human thought and action.

Recently, some theorists have become increasingly bold in asserting that consciousness is an epiphenomenon with no impact on behavior. We have been skeptical that such a complex, advanced phenomenon as human conscious thought would evolve without conferring vital advantages. Inspired by this skepticism, we conducted an extensive review of experimental evidence for the conscious causation of behavior (Baumeister et al. 2011). We searched for experiments in which the independent variable was a conscious thought or other conscious event and the dependent variable was overt behavior. By the logic of experimental design, such studies prove causation. We found a wide assortment, leading to our conclusion that the evidence for conscious causation was extensive, diverse, and undeniable.

Still, none of the evidence we found ruled out important, even essential contributions by unconscious processes. We speculated that there may well be no human behaviors that are produced entirely by consciousness (and likewise few produced entirely by unconscious processes). If all behavior is indeed produced by both conscious and unconscious processes, then it is imperative to understand both types and how they interact.

In fact, we think that conscious processes are themselves produced in part by unconscious processes. Baumeister and Masicampo (2010) concluded that consciousness is best considered a place where the unconscious constructs meaningful sequences of thought. This is linked to evidence that, for example, the unconscious processes single words but not sentences and paragraphs (see Baars 2002). Likewise, logical reasoning deteriorates sharply when consciousness is preoccupied and improves when engaged (DeWall et al. 2008). Logical reasoning requires putting together complex sequences of ideas while ruling out other possible sequences, and that may be too big a job to do effectively without consciousness.

What else is consciousness good for? We are wary of making assertions that something absolutely cannot be done unconsciously – but perhaps that is not necessary. After all, the capacity for conscious thought would have been favored by natural selection simply on the basis of doing something better or more thoroughly than unconscious processes, even if the unconscious could occasionally do something along the same lines well enough to produce an experimental finding. For example, one recent paper has proposed that the unconscious can do some arithmetic (Sklar et al. 2012). Even if this finding could measure up to the methodological standards proposed by N&S, we think that is hardly a reason to dismiss the usefulness of conscious thought for mathematical work. Does anyone seriously think that a student could pass a college math test without conscious thought?

Key themes from our survey of experimental findings on conscious causation (Baumeister et al. 2011) included the following: Conscious thoughts integrate across time. That is, conscious thought permits the deliberate combining of past and future into causing present behavior, as well as helping present cognitions to cause future behavior, and probably other combinations. Most animals live largely in the present, whereas the stupendous success of the human race has benefited immensely from integrating across time – for which conscious thought probably deserves much of the credit.

Conscious thought also helps translate abstract principles into specific behaviors. Humankind has benefited from moral principles, legal rules, economic calculations, application of scientific and mathematical principles, and other sorts of general, abstract understandings. The unconscious may be effective at processing highly specific stimuli and responses, but without conscious thought, it may lose most of the benefits of abstract principles for guiding behavior.

We also found that conscious causation of behavior was typically found in situations involving multiple possibilities, such as for

deliberating among multiple possible courses of action by considering their likely consequences. Consciousness enables the person to mentally simulate nonpresent realities, including possible sequences of future events. In our view, the jury is still out on whether the act of choosing is conscious or unconscious – but most decisions will be considerably more effective insofar as one uses conscious thought to ponder what each possible action will produce and the desirability of downstream consequences.

Last, and perhaps most important, consciousness is highly useful for communication. We have not heard even the most assertive critics of conscious thought claim that a person could carry on a conversation unconsciously. Indeed, we think that the evolutionarily decisive advantages of conscious thought are not to be found in private, solipsistic ratiocination but rather in its contribution to communication (Baumeister & Masicampo 2010). Humankind's biological strategy for surviving and reproducing has been centrally based on sharing information and coordinating joint performance. Although the principle that thinking is for doing has been widely considered sacrosanct since first asserted by William James (1890), we propose that a viable partial alternative is that conscious thinking is for talking.

In sum, we applaud N&S for pushing the field forward. Conscious thought is a vital part of human life. We think the view that humans could operate effectively without conscious thought will soon be regarded as quaint and naïve.

The problem of consciousness in habitual decision making

doi:10.1017/S0140525X13000642

Javier Bernacer,^a Gloria Balderas,^a Ivan Martinez-Valbuena,^a Maria A. Pastor,^{a,b} and Jose Ignacio Murillo^a

^aMind–Brain Group, Institute for Culture and Society (ICS), University of Navarra, 31009 Pamplona, Navarra, Spain; ^bNeuroimaging Laboratory, Neurosciences Section, Center for Applied Medical Research, University of Navarra, 31008 Pamplona, Navarra, Spain.

jbernacer@unav.es gcbalderas@gmail.com

imartinez.18@alumni.unav.es mapastor@unav.es

jimurillo@unav.es

http://www.unav.es/centro/cultura-y-sociedad/mind_brain

Abstract: Newell & Shanks (N&S) carry out an extremely sharp and static distinction between conscious and unconscious decisions, ignoring a process that dynamically transfers decisions and actions between the conscious and unconscious domains of the mind: habitual decision making. We propose a new categorisation and discuss the main characteristics of this process from a philosophical and neuroscientific perspective.

Newell & Shanks (N&S) establish the elements of a decision by using the lens model (Brunswik 1952). According to their framework, decisions can be *either* conscious or unconscious: In the former, all five stages of the lens model are supervised by consciousness, whereas in the latter at least one of the five elements is unconsciously performed. In our opinion, this is an extremely sharp distinction that leaves out of the picture aspects of a crucial importance in action selection, such as habits or habitual decision making.

Human agents make many decisions every day, some of which are fully unnoticed. Considering the role of consciousness in the course of decision making, we propose three categories in which a decision can be included: (1) conscious decisions, (2) retrospective attributions to unconscious behaviour, and (3) non-conscious but controlled decisions. The first type refers to deliberative decisions, which are made when facing a problem that requires a high cognitive load, especially involving a novel situation. Retrospective attributions are actions performed under low or non-existent levels of consciousness, and whose meaning is attributed