

perceptual templates, attempting to achieve a 'best match' with the incoming data. . . . [T]he result is globally broadcast as part of the perceptual state itself" (sect. 2, para. 10).

Although this is certainly plausible, Carruthers neglects the fact that similar grounds are available for the claim that other propositional attitudes are broadcast as well. Indeed, given his concession that global workspace theories allow for the broadcasting of at least one propositional attitude, one wonders why he assumes that such theories would *not* allow for the broadcasting of all the rest. On the face of it, the claim that *all* propositional attitudes can be globally broadcast has much going for it. Intentions, for instance, routinely recruit a wide array of cognitive resources, as do the conceptual-intentional aspects of emotions like fear and anger (e.g., *that* one is being attacked). Why not count these as instances of global broadcasting? Carruthers does not say.

Pending further argument, we should assume, *pace* Carruthers, that global workspace theory *does* allow for the broadcasting of all propositional attitudes. If so, then whatever we say about first-person access to sensory states, we should say the same about first-person access to propositional attitudes.

Do these considerations support the view that the mindreading system has direct, non-interpretative access to *all* mental states, both propositional and sensory? Not if one also rejects Carruthers' assumption that globally broadcast states are *ipso facto* available to the mindreading system in a *non-interpretative* fashion. Below, I explore grounds for adopting the strong version of MPM, according to which self-attribution is interpretative in the case of all mental states.

Interpretation takes place by deploying a propositional attitude that emerges from a background of theoretical commitments. Consequently, the cost of embarking on an interpretative venture is the possibility of partial misconstrual or wholesale error. These characteristics of interpretative activity fit well with Carruthers' usage of the term "interpretative," as applied to mechanisms of self-attribution.

As Rosenthal (2005) has argued, self-attribution is a matter of tokening potentially erroneous, theoretically loaded propositional attitudes – occurrent higher-order thoughts (HOTs). On this view, confabulation and error occur even with regard to sensory states. Dental fear, for instance, is a phenomenon in which dental patients under the drill mistake fear, anxiety, and a sensation of vibration for pain in a fully anaesthetized or nerveless tooth – a compelling demonstration that HOTs need not be veridical.

Nevertheless, judged on independent grounds, self-attributions of sensory states are often relatively accurate. Doubtless, this consideration compels theorists to posit a reliable monitoring mechanism, such as Carruthers' mindreading system. But, as Rosenthal points out, simply positing such a mechanism amounts to no more than *stipulating* a solution to the problem of explaining the frequent accuracy of HOTs about sensory states. An explanatory account of the mechanism's accuracy is not provided.

Extending Sellars's (1956/1997) treatment, Rosenthal argues that HOTs concerning sensory states arise as a result of a creature's reflection on cases in which its perceptual judgments are mistaken. The creature formulates a rudimentary theory, in effect positing qualitative sensory states as the causes of non-veridical perceptual judgments. Against the background of such a theory, the creature is disposed, for instance, to construe itself as having a sensation of red when perceiving a red object.

Carruthers gives no grounds for rejecting this alternative and appealing picture. Global broadcast theory does not, by itself, settle the issue, for it is consistent with the claim that the mindreading system relies on a tacit theory in interpretatively self-ascribing sensory states. Nor does the data from autistic children disconfirm Rosenthal's view, which allows that even non-linguistic, cognitively unsophisticated creatures may come to have

HOTs concerning their sensory states. All that is required is that such creatures take note of their perceptual errors and account for them.

ACKNOWLEDGMENT

I am grateful to David Rosenthal for helpful comments on an earlier draft.

Introspection and interpretation: Dichotomy or continuum?

doi:10.1017/S0140525X09000764

Richard E. Petty^a and Pablo Briñol^b

^aDepartment of Psychology, Ohio State University, Columbus, OH 43210;

^bDepartamento de Psicología Social, Universidad Autónoma de Madrid, 28049 Madrid, Spain.

petty.1@osu.edu www.psy.ohio-state.edu/petty

pablo.brinol@uam.es www.psy.ohio-state.edu/gap/Pablo/pablo.htm

Abstract: Judgments vary in the extent to which they are based on interpretation versus relatively direct access to mental contents. That is, a judgment might require a trivial amount of interpretation (e.g., translating one's immediately accessible "inner speech") or a rather substantial amount of confabulation. Recognizing this continuum of interpretation underlying judgment could be more fruitful than debating a categorical introspection versus interpretation distinction.

Some prior authors have noted that people have no unique access to *why* they believe what they believe (e.g., Nisbett & Wilson 1977). Others have gone a step further and postulated that people do not know their own attitudes (e.g., I like ice-cream) but must construct them when needed from other available information that they either retrieve from memory (e.g., ice cream tastes good) or extract from the immediate context (e.g., Schwarz & Bohner 2000). Carruthers takes this "constructivist" position to the ultimate extreme by arguing that people have no direct access to any attitudes *or* relevant beliefs. According to this view, introspection does not exist, and is merely an illusion. Furthermore, he provides many examples where people either clearly or plausibly are confabulating when they express what they believe. In his view, at best individuals only know what they feel and perceive, not what they think. In our view, it is not clear why an intelligent organism would have evolved to have direct access to its feelings and perceptions but not its cognitions.

Nevertheless, Carruthers has an important point. Whenever someone expresses a belief or has a thought, some degree of interpretation likely is involved, if only to understand the meaning of the "inner speech" in which the thought is expressed. Thus, if a person has a positive reaction to some stimulus (ice-cream), this can be translated into "it's good" (Fazio 1985). Or even if the word "good" immediately springs to mind, the meaning of the word must be understood by the self if an internal thought, or by others if expressed. However, this very minimal form of "interpretation" is very different from the kind of interpretation involved in most of the examples of confabulation provided by Carruthers. Indeed, we argue that it may not be wise to think of introspection and interpretation as dichotomous categories in which to place any given judgment. Rather, there are various degrees of interpretation. At the low end of the interpretation continuum, judgments are introspection-like in that they involve at most some trivial interpretation. At the other end of the continuum, the judgment is totally confabulated from external sources also available to outside observers.

Although dichotomous categories can be useful in understanding some phenomena, as illustrated by the distinction between primary and secondary (meta-) cognition, we believe that it is not conducive to understanding human information processing

to simply lump all judgments into the same overarching “interpretation” category and stop there. This is because putting all judgments into the same category might suggest that there are no meaningful differences *within* the category. In contrast to lumping all judgments together into one interpretation category, we espouse a continuum view in which people express beliefs based on very little interpretation in some cases but based on extensive confabulation in others. We further argue that differences in the *degree* of interpretation are meaningful.

Previous research on psychological elaboration provides one instance of the usefulness of the continuum view. The term “elaboration” is used in social psychology to describe that people add something of their own to the specific information provided, for example, in a persuasive communication. In the domain of social judgment, variations in elaboration are consequential. For example, when people are relatively unmotivated or unable to think, they are more likely to rely on immediately accessible information that originates either internally (one’s attitude) or externally (e.g., the attractiveness of the message source). When people are more motivated and able to think, then these initial reactions and the judgments that follow from them can be overridden by more complete interpretative analyses. Furthermore, judgments based on high levels of elaboration are more consequential than those based on low levels (Petty & Cacioppo 1986).

Viewing interpretation as a continuum has a number of implications. Most obviously, it means that interpretation can go from zero (i.e., introspection) to extensive. More interestingly, the continuum view suggests that the point on the continuum that corresponds to minimal or trivial interpretation has more in common with zero interpretation than it does with extensive interpretation. One can draw an analogy to a distinction that attitude theorists used to favor between attitude formation versus attitude change. Different mechanisms of influence were thought to be operative depending on whether a person had an existing attitude or did not (a categorical view). Today, it is more common to think of attitudes as falling along a continuum such that they can vary in how accessible they are or upon how much knowledge they are based. An attitude formation situation would be present when a person has no prior attitude. But, a continuum approach to attitudes suggests that a person who has an attitude that is difficult to bring to mind and based on little information (Person B) has more in common with an individual who has no attitude (Person A) than a person who has an attitude that comes to mind spontaneously and is based on much knowledge (Person C). That is, the first two individuals – A and B – are more similar to each other in psychologically relevant ways than they are to C, despite the fact that a dichotomous approach places Person B in a different category from A and in the same category as C. So too is it the case that a judgment based on minimal interpretation (B) is closer to a judgment based on no interpretation (A) than it is to a judgment based on extensive interpretation (C; see Fig. 1).

In sum, we conclude that an all-or-none frame regarding the existence of introspection may not be the best way to make the most progress in understanding social judgment. Instead, drawing from the literature on elaboration and attitude strength, we suggest that it might be more fruitful to approach

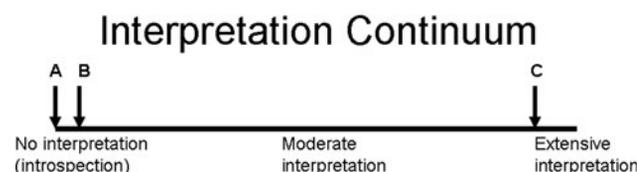


Figure 1 (Petty & Briñol). Continuum of extent of interpretation underlying judgment

interpretation as a continuum where the low end is anchored at introspection. That is, sometimes interpretation can be quite minimal as when people recall their birth-date or liking of a favorite book. At other times, interpretation can be quite extensive, such as when there is either nothing relevant to recall or one’s interpretation totally overwhelms any mental content introspected.

Overlooking metacognitive experience

doi:10.1017/S0140525X09000776

Joëlle Proust

Department of Cognitive Studies, Ecole Normale Supérieure, and Institut Jean-Nicod, EHESS and ENS, 75005 Paris, France.

jproust@ehess.fr

<http://joelleproust.hautefort.com>

Abstract: Peter Carruthers correctly claims that metacognition in humans may involve self-directed interpretations (i.e., may use the conceptual interpretative resources of mindreading). He fails to show, however, that metacognition cannot rely exclusively on subjective experience. Focusing on self-directed mindreading can only bypass evolutionary considerations and obscure important functional differences.

Carruthers’ main goal is to show that metacognition is a form of self-directed interpretation, akin to other-directed mindreading. Introspection, he claims, defined as “any reliable method for forming beliefs about one’s own mental states that is *not* self-interpretative and that differs in *kind* from the ways in which we form beliefs about the mental states of other people” (sect. 1.4, para. 3, emphasis in the original), is not needed to have access to one’s own mental attitudes. One can agree with the author that metacognition in humans *may* involve self-directed interpretations (i.e., *may* use the conceptual interpretative resources of mindreading), without accepting the stronger claim that metacognition can *never* be based on “introspection.”

In cognitive science, “metacognition” refers to the capacity through which a subject can evaluate the feasibility or completion of a given mental goal (such as learning a maze, or discriminating a signal) in a given episode (Koriat et al. 2006). In Carruthers’ use, however, metacognition *refers to* first-person metarepresentation of one’s own mental states; as a result, the theoretical possibility that metacognition might operate in a *different* representational format cannot be raised (Proust, in press b). Revising the meaning of a functional term such as “metacognition” is a bold strategy. It generally seems more adequate to leave it an open empirical matter whether a capacity of type A (reading one’s own mind) or type B (evaluating one’s cognitive dispositions) is engaged in a particular task. A revision is deemed necessary, according to Carruthers, because “B” capacities in fact always involve self-directed mindreading; therefore apparent contrary cases (self-evaluation in non-mindreading animals) either (1) are simply instances of first-order types of learning, and/or (2) are capacities “too weak to be of any interest” (Carruthers 2008b, p. 62; cf. target article, sects. 5.1 and 9).

Two methodological problems, however, hamper the discussion so conceived. First, it is quite plausible that, in *human* forms of metacognition (as instantiated in speech production, metamemory, etc.), judgments of self-attribution *re-describe* elements of metacognitive experience. Metacognitive feelings might, on this view, represent subjective uncertainty and guide noetic decision-making, without needing to involve a conceptual interpretative process. What needs to be discussed, in order to establish the superiority of model 4, is whether or not subjects can rely on dedicated feelings alone to monitor their ongoing cognitive activity.

A second, related problem is that Carruthers’ discussion conflates two domains of self-control, namely, the control of one’s