believe identity-bolstering fake news, and to generate politicized conspiracy theories. These examples highlight the critical role of social identity in rationalization:

- *Resist evidence.* People often discount or rationalize evidence that contradicts their firmly held political beliefs or party affiliation. For instance, people are less likely to update their political views in the face of counterevidence compared to their non-political views (Kaplan et al. 2016). This belief resistance was associated with activity in the prefrontal cortex – suggesting a role for motivated reasoning or rationalization. In some cases, exposure to opinions from political out-group members can even backfire – making people more entrenched in their political beliefs than before (Bail et al. 2018).
- *Rationalize lies.* Cushman argues that rationalization allows people to translate gut instincts into rational thoughts that "represent[s] true properties of the world." However, people readily rationalize false information when it is propagated by party elites and aligns with their political identity For example, all Trump-branded hats are made in the United States, but when Clinton supporters were told to imagine that Trump would make his merchandise outside the United States if it were cheaper to do so, they felt it would be less unethical to lie that this merchandise *was* made outside the United States, and that political elites who espouse these lies deserve less punishment (Effron 2018).
- *Believe fake news:* People also rationalize fake news that is positive about one's in-group or negative about one's out-group. For example, Democrats were more likely to believe negative fake news about Republican politicians than negative fake news about Democratic politicians, and vice versa for Republicans (Pereira et al. 2019). People are typically motivated to hold true beliefs, but in the case of identity-bolstering fake news, it can be more beneficial to rationalize these false beliefs as true. In believing them, people can share similar beliefs with in-group members and maintain positive beliefs about the group.
- *Generate conspiracy theories.* Conspiracy theories connect different, unrelated, and inconsistent events in a way that seems meaningful and rational. As such, conspiracy theories can help uphold a positive group-identity under the guise of rationality. For instance, some scholars argue that *conspiracy theories are for losers,* such that the loss of political power increases conspiracy theory beliefs (Uscinski & Parent 2014). Indeed, prior to the 2012 U.S. presidential election, Republicans and Democrats were similarly likely to expect electoral fraud. However, after President Obama was re-elected, Republicans were more likely to believe that electoral fraud had occurred (Edelson et al. 2017). Reducing the their loss to a conspiracy allowed Republicans to rationalize and uphold positive beliefs about their in-group.

This is a sample from a large literature exploring the social function of rationalization. Factors that increase identification with political parties or movements can increase the value of rationalization given that it may help people remain in good standing with fellow group members (see Van Bavel & Pereira 2018). Furthermore, rationalizing actions of in-group elites can reduce accountability for harmful behavior and create conflict with out-group members, thus increasing polarization. At the same time, polarization can increase commitment and identification with one's in-group, thereby motivating rationalization. Thus, aspects of the intergroup context, like polarization, can both amplify rationalization and result from group-based rationalization.

Understanding the role of social identity in rationalization is not only critical for understanding the function(s) of this psychological process, but also clarifying when and why features of the context will elicit and result from rationalization. For instance, situations that increase the salience of identities or the norms associated with those identities will impact rationalization. These forms of rationalization not only help an individual maintain or increase their standing within the group (which can promote well-being and survival), but also ensure that the group maintains cohesion during intergroup competition.

The rationale of rationalization

Walter Veit[a], Joe Dewhurst[b], Krzysztof Dołęga[c], Max Jones[a], Shaun Stanley[a], Keith Frankish[d] and Daniel C. Dennett[e]

[a]Department of Philosophy, University of Bristol, Bristol, BS8 1TH, United Kingdom; [b]MCMP, LMU Munich, 80539 Munich, Germany; [c]Institut für Philosophie 2, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum; [d]Department of Philosophy, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom and [e]Center for Cognitive Studies, Tufts University, Medford, MA 02155.

wrwveit@gmail.com https://walterveit.com/
joseph.e.dewhurst@gmail.com https://joedewhurst.weebly.com/
krzysztof.dolega@rub.de https://www.krysdolega.xyz/
Max.Jones@bristol.ac.uk http://www.maxjonesphilosophy.com/
shaun.stanley@bristol.ac.uk8
k.frankish@sheffield.ac.uk https://www.keithfrankish.com/
Daniel.Dennett@tufts.edu https://ase.tufts.edu/cogstud/dennett/

doi:10.1017/S0140525X19002164, e53

Abstract

While we agree in broad strokes with the characterisation of rationalization as a "useful fiction," we think that Fiery Cushman's claim remains ambiguous in two crucial respects: (1) the reality of beliefs and desires, that is, the fictional status of folk-psychological entities and (2) the degree to which they should be understood as useful. Our aim is to clarify both points and explicate the rationale of rationalization.

Post hoc rationalization, that is, retrospectively attributing or constructing "hidden" beliefs and desires inferred from how one has behaved in the past, has traditionally been seen to threaten the idea that humans are "rational," since it happens subsequent to the process under consideration. If the relevant mental states that are supposed to rationalise an action only come into existence after the action has occurred, then they cannot be treated as the cause of that action. However, Cushman argues that a post hoc process of this kind can still be seen as "rational" in the sense that it constructs new beliefs and desires that both serve a useful function and track some underlying adaptive rationales that have shaped the behaviour being rationalised. Rationalization, according to Cushman, is supposed to be a "useful fiction." We think that this proposal invites two serious ambiguities: first, to do with the ontological status of the mental states that are the outputs

of rationalization (i.e., folk-psychological states like beliefs and desires) and, second, to do with the degree to which they should be understood as useful and representative. We will address each ambiguity in turn, using our resolution of the latter to help resolve the former.

Throughout his article, Cushman seems to assume a fairly robust understanding of what beliefs and desires are, framing them as functionally discrete internal states with determinate contents. He is committed to the idea that there is a crucial distinction between "real" reasoning processes, which involve operations on beliefs and desires, and the fictional ones produced by rationalization, which don't involve any such operations. Rationalization, on his account, seems to play the role of a process of self-interpretation in which one authors fictions about the causes of one's own behaviour. Drawing these distinctions might not be as easy as Cushman suggests, if there is no principled "dividing line between *genuine* belief-talk or agent-talk and mere *as if* belief-talk and agent-talk" (Dennett 2011, p. 481). Indeed, the lack of such a dividing line similarly arises for agential descriptions or "rationalizations" in evolutionary biology (see Dennett 2019; Okasha 2018; Tarnita 2017; Veit 2019). Without such a dividing line, however, it is unclear what the ontological status of beliefs and desires is supposed to be. If Cushman were to deny that there are anything at all like beliefs and desires prior to the rationalization process, making the folk-psychological states produced by this process entirely fictional, he would fall close to eliminative materialists such as Paul and Patricia Churchland (Churchland 1981; 1986). We do not think that Cushman would like to endorse this option, as he seems quite committed to the existence of beliefs and desires. The other option, then, and this is a move we recommend for Cushman, is to commit to the existence of some sort of proto-mental states prior to the rationalization process, in which case we think it is unclear in what sense the output of the rationalization process also constitute fictional entities. Of course, the rationalization process might influence or replace these proto-mental states via a narrative process that we could call fictional, but it is no longer the mental states themselves that are fictions, rather the process that produces them.

This brings us to the second ambiguity: In what sense can fictional mental states (or processes) be understood as useful? Cushman clarifies that these fictions can be useful even when they are not "perfectly accurate representations" by appealing to Dennett's (1987) "intentional stance," according to which the attribution of beliefs and desires are understood as nothing more than a way of tracking observable patterns in behaviour (or the categorical bases of those patterns) and have no further ontological status *inside* the system. However, this comparison reveals a tension in his dual conception of folk-psychological states. Dennett's intentional stance assumes that habit, instinct, norms, and so on, may all support rational patterns of behaviour, and that this is all that is needed for a system to manifest genuine beliefs and desires. It is true that these processes support rational responses that make it worth extracting information from them via rationalization (i.e., by adopting the intentional stance) and then re-presenting this information in a rich belief/desire format. Reformatted in this way, beliefs and desires take the form of the linguistic utterances that Dennett (1987) originally called "opinions" and Frankish (2004) has more recently called "superbeliefs." For us, richness is a matter of having a discrete representational vehicle, such as that provided by natural language, but it is not clear that this is what Cushman has in mind when he talks about beliefs and desires.

As we see it, there are two broad ways to achieve such a rich conception of belief, either internal or external. On the internal conception, that is, traditional computationalism, this vehicle is a neural one, and beliefs are formed and processed at a subpersonal level. On the external conception, the vehicle is natural language, and beliefs are formed and manipulated at a personal level by agents themselves, as a way of describing and regulating their own and others' behaviour. Forming a rich belief, that is, an opinion or superbelief, is like adopting a policy or making a bet on truth – we commit to taking a sentence as an expression of truth and regulate our other utterances and commitments accordingly. Cushman seems to espouse a version of the former interpretation, but we think that the latter interpretation is to be preferred, as it can help to resolve the two ambiguities outlined above.

Once this external approach is adopted, the sense in which rationalization is *fictional* becomes clear: It involves the construction of a narrative that is strictly false with regard to the underlying mechanisms, but nonetheless captures real patterns in the behaviour generated by those mechanisms. We propose to interpret rationalization as the process of taking the austere "proto-beliefs" manifested in behaviour and transforming them into superbeliefs or opinions (i.e., rich, linguistically formatted beliefs and desires) via the application of the intentional stance to one's own behaviour. Taking this can help to resolve the ambiguities described above, provided that Cushman is willing to adopt this distinction between the austere beliefs that are implicit in all (seemingly) intelligent behaviour, and the explicit, linguistically mediated beliefs that are the outcome of the rationalization process. The latter could be seen as fictional, in the sense that they only came about as the result of a story that we tell about our own behaviour, and yet they are also real, in the sense that they do accurately capture (and help to track) our behaviour (even if they do not accurately describe the processes underlying that behaviour). By coming to be explicitly represented in natural language, expressing normative commitments, they can also indirectly influence our future behaviour. In short, we think rationalization should be treated as the reverse engineering of what Dennett (2017) has called "free-floating rationales," that is, instinctive behavioural patterns, like avoiding snakes or heights, that are not explicitly encoded but nonetheless make rational sense. Similarly, the underlying *reasons* that are implicit in our behaviour can be inferred (or rather uncovered) via rationalizations, which can then lead to further behavioural improvements by engaging in explicit rational deliberation. This is the rationale of rationalization.

Hard domains, biased rationalizations, and unanswered empirical questions

Stephen E. Weinberg[a] and Jonathan M. Weinberg[b,1]

[a]Department of Public Administration, Rockefeller College of Public Affairs and Policy, University at Albany, State University of New York, Albany, NY 12222 and [b]Department of Philosophy, University of Arizona, Tucson, AZ 85721-0027.
sweinberg@albany.edu jmweinberg@email.arizona.edu
https://www.albany.edu/rockefeller/faculty/stephen-weinberg

doi:10.1017/S0140525X1900222X, e54