and mutual endorsement of the standards. In the face of deviations from standards, people have to express condemnation in order to make the violated standard salient to all group members (Feinberg 1965; Durkheim 1893). In addition, the expression of blame for norm violations demonstrates that group members care about the norms and the group members protected by those norms. Finally, observers blame norm violators to distance themselves from the deed and avoid being associated with such misdeeds. Thus, in some sense the observers show agency when they blame and praise others' behavior because it expresses their values (usually, shared values). They may even express their values without caring too much for actual responsibility of the actors (i.e., they may not go further than differentiating between coerced and uncoerced behavior).

Moreover, we argue that the assignment of blame or praise for misdeeds also affects the actors' agency. Public condemnation indicates, claims, or even fosters group members' exercise of agency. As observers attribute responsibility to the actors, the actors may also perceive themselves as having agency (or an illusion of agency?). For example, children's agency develops by the guidance of sanctions. Agency may be considered an *ability* (that one could learn) instead of a *habit*. Habits denote what people are accustomed to do, whereas abilities include a normative component that denotes what would count as a correct or incorrect thing to do (Millikan 2000). This normative component specifies when we sometimes succeed in expressing our values and when we fail to express them. As mentioned above, praise and blame direct us thereby in the standard's (valued) direction. In contrast, habits could go in any direction, as they are not necessarily corrected by values. Moreover, by such development of ability over time (i.e., agency-training), we become more reliable in expressing our values in particular situations and apply them to more diverse situations.

As an additional mechanism, we suggest that reminders of our responsibility, such as blaming and praising of certain behaviors, activate the concept of personal agency. Activated concepts also tend to produce concept-related behavior (e.g., the belief that one excels in math enhances math performance, Miller et al. 1975). Activated concepts also change cognitive processing characteristics that lead to the enactment of these concepts (Sassenberg et al. 2017). Accordingly, actors who are held responsible may activate their concept of "being responsible." Thus, before acting, they may think twice, activate their main values, and take precautions to make sure that their behavior conforms to their values. Such a reflection of personal values in turn may lead to a stronger connection of these standards in their cognitive system; they may identify with them and thereby behave more in accordance to them. This is also a social process: it not only involves solitary thinking but also social negotiation and training in justifying behavior in the face of others. This may reward careful action, so that people may arrange their environment in order to avoid known "defeaters" (e.g., temptations). Moreover, being held responsible indicates being watched. This enhances objective self-awareness and thereby a person's own standards become more salient.

The social shaping of agency and responsibility may not always work out completely. Some people may be hard to train or unwilling to develop stable "virtues" (i.e., habits to act according to their own and commonly shared standards). However, this may be irrelevant, as others will still hold them responsible (even if this cannot apply literally) and punish them (e.g., go for incapacitation as a last resort). In addition, people may not want to wait until repeated misdeeds manifest the "negative" values of the actor. There may be an asymmetry in that many positive deeds are necessary to manifest positive values of people, whereas one negative deed can be enough to reveal the negative value of an actor. The extremity of the deed may itself be a clear indicator for moral responsibility (Pauer-Studer & Velleman 2011). In such cases, where the social shaping of individual agency or responsibility may be impossible or come too late, the actor can only be made incapable. However, the general practice of collaboratively shaping agency may not be threatened by this because these examples remain exceptions.

In short, the emergence of agency and responsibility is a social process. Talking to others (including blaming and praising) is a particularly efficient way to develop one's own agency and help others become responsible actors.

Grounding responsibility in something (more) solid

doi:10.1017/S0140525X17000711, e47

William Hirstein and Katrina Sifferd

Department of Philosophy, Elmhurst College, Elmhurst, IL 60126.

williamh@elmhurst.edu **sifferdk@elmhurst.edu**

Abstract: The cases that Doris chronicles of confabulation are similar to perceptual illusions in that, while they show the interstices of our perceptual or cognitive system, they fail to establish that our everyday perception or cognition is not for the most part correct. Doris's account in general lacks the resources to make synchronic assessments of responsibility, partially because it fails to make use of knowledge now available to us about what is happening in the brains of agents.

Our commentary on Doris's significant book focuses on three areas: (1) Doris's claim that cases of self-ignorance, such as confabulation, are common enough to negate our own judgments of why we did things; (2) Doris's inability to give a good account of synchronic assessments of responsibility; and (3) the disconnect between Doris's account and scientific accounts of human thought and behavior.

Self-ignorance. Doris says that human beings are "afflicted with a remarkable degree of self-ignorance" (précis abstract). But while we certainly at times show self-ignorance, there is no absolute metric that allows us to assess the exact degree of our ignorance compared to our self-knowledge. This opens the possibility for researchers, who feed on a steady diet of examples of ignorance, to overestimate its degree. We need to leave open, for example, the possibility that we are dealing not with phenomena that afflict everyone, but with phenomena that only afflict a minority of people, or even a certain personality type. The scope of Doris's skepticism is also broader than it might appear. One sign that we might be overestimating the amounts of ignorance and error is that we have not been moved to enact major changes in folk-psychology to remove dependence on our capacity for self-knowledge. Doris's view seems to commit us not only to being "routinely mistaken" (précis abstract), but also not ever noticing that we are, and attempting to correct it. Doris seems to be neglecting all those times we *aren't* buffoons.

A comparison with the case of visual perception is illuminating. Even though cognitive scientists have cataloged perhaps hundreds of visual illusions that reveal the seams and flaws of our visual system, the vast majority of our visual perceptions during the day are veridical and serve us quite well. Consider our abilities to visually identify one another. Certainly there are many ways in which the brain systems that achieve this miracle can fail, leading to odd syndromes like prosopagnosia. In the everyday sphere, we have all experienced cases in which we visually misidentified someone. But taken against the overwhelming percentage of correct identifications we make so effortlessly and frequently, these misperceptions are rare. This high rate of effectiveness is due to good equipment.

We think serious cases of ignorance or mistaken self-knowledge are somewhat rare because they typically involve errors at two levels. First, a mistaken impression is created. For instance, it occurs to me that I don't really have to pay back that loan from my friend because he seems to be wealthy, when I would just

Downloaded from https://www.cambridge.org/core. Cambridge University Press, on 28 Mar 2018 at 16:01:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0140525X17000711