Intentions: Philosophical and Empirical Issues

Markus Schlosser · Fabio Paglieri

This *topos* is focused on intentions, with an emphasis on integrating philosophical analysis and empirical findings. Theorizing about human action has a long history in philosophy, and the nature of intention and intentional action has received a lot of attention in recent analytic philosophy. At the same time, intentional action has become an empirically studied phenomenon in psychology, cognitive neuroscience, artificial intelligence, and robotics. Many results obtained in these areas have been incorporated within the current philosophical debate, while at the same time scientists have often adopted in their experiments and models philosophical assumptions on the nature of intention and intentional action. As a result, the study of intentions is nowadays a thriving enterprise, where both conceptual and empirical issues are discussed in a dialogue across disciplines. This is well reflected in the selection of papers published here.

Davide Rigoni and Marcel Brass discuss the social and neural consequences of disbelieving in free will. Contemporary neuroscience enables the experimental investigation of complex psychological functions related to free will, such as conscious intention, decision-making and self-control. The findings of this research have attracted a lot of media coverage, with frequent claims to the effect that free will is nothing more than an illusion. Rigoni and Brass ask whether, why, and how such neuroscientific findings influence our everyday belief in free will. Based on an extensive review of the literature in experimental philosophy, social psychology and cognitive neuroscience, they

show that inducing disbelief in free will has an impact on folk psychology, social behavior and intentional action.

Ariel Furstenberg argues for the existence of non-executed unconscious proximal intentions, i.e., unconscious proximal intentions to act that do not result in overt movement. He first presents a conceptual framework that accounts for the phenomenon of non-executed proximal intention and the related phenomenon of change in proximal intention. Then he turns to empirical findings, claiming that a specific EEG signal could provide a neural correlate of a non-executed proximal intention, thus justifying usage of the concept of "intention" even for mental states that never result in overt action.

The question of how to account for intentions in the absence of overt action is also central in Zoe Drayson's discussion of mental agency in post-coma patients. She critically reviews recent findings which suggest that intentional mental action may be reliably revealed by a certain pattern of neural activity, thus allowing to attribute conscious awareness to patients in vegetative states, with important legal and ethical consequences. However, Drayson identifies two key weaknesses in this so-called 'argument from volition': first, while the neuroimaging data may provide evidence for the existence of certain mental events, it is neutral with regard to whether these mental events constitute mental actions; second, it is difficult to see how one could set up a neuroimaging task that would enable us to make the required discrimination.

Élisabeth Pacherie tackles a foundational question in the philosophy of action: How do conscious intentions relate to actions? She first presents the traditional philosophical view of the structure of agency, in which conscious intentions are the causes of actions, as well as two important worries raised by recent empirical findings: skepticism about the role of consciousness in the causation

M. Schlosser (✉) · F. Paglieri
School of Philosophy, University College Dublin, Dublin, Ireland
e-mail: markus.schlosser@ucd.ie

of action and skepticism about the assumption that intentions causally *initiate* movements (both fueled mainly by the famous Libet experiment and by Wegner's work on the 'illusion of conscious will'). She then argues that both the traditional view and some of the objections against it rest in part on an over-simplified conception of the structure of agency. If we take into account the hierarchical nature of intentions and the role of control processes, we can construe the relation between intentions and actions differently, and we can save conscious agency from the empirical challenge.

Lilian O'Brien's contribution is about the so called 'simple view' of intentional action: the claim that an intention to A is necessary if I am to A intentionally. It is widely accepted that Michael Bratman's thought experiment on mutually exclusive planning undermines the simple view. O'Brien does not aim to defend the simple view, but she argues that close examination of Bratman's case reveals a dilemma: either the mental state involved in the case is an intention to A after all, or the case does not involve an intentional action—either way, the simple view is not undermined. For O'Brien, the real question raised by Bratman's scenario is how we can rationally intend mutually exclusive ends. Her solution to this puzzle rests on considerations concerning the subjective authority of intentions, according to which the agent is not always criticizable for failing to perform an intended action.

Till Vierkant discusses the ability to stop ourselves from giving into temptation (self-control), and challenges the intuition that doing so by "sheer willpower" is importantly different from using external props that simply prevent one from undermining previous resolutions (e.g., Ulysses tying himself to the mast of his ship to avoid the lure of the Sirens). In contrast, Vierkant argues that the kind of agentive control over our own mental states involved in willpower is not essentially different from the kind of control required to use the external environment to help ourselves resisting temptations: in particular, both forms of self-control are equally indirect, work by suspending rational evaluation rather than by enhancing it, and grant the same (modest) level of behavioral flexibility. If this is correct, then it is easy to see that there are even stronger grounds to think that the will is extended than there are for thinking that cognition is extended.

The opposite of willpower, that is, weakness of the will or *akrasia*, is the main topic of Gregory Strom's contribution. He starts with an argument against Davidson's classic account of *akrasia*, thus motivating a new analysis of the phenomenon, based on a distinction between different ways of having practical reasons. On this view, akratic agents have existential knowledge that there is some decisive practical reason to act otherwise, but they do not really know what that reason is. Without knowing what that

reason is, an agent can act in accord with it only in a way that deviates from the conditions under which her action might manifest that reason. Then Strom uses this assumption to argue that practical rational excellence is not just a matter of doing what one has reason to do, but also a matter of safely transforming practical knowledge of reasons into actions.

In Marco Mazzone's paper, the focus shifts to an alleged parallelism between linguistic behavior and intentional conduct in general, regarding the possible existence of a generative system for intentional actions. Structural similarities between language production and action planning and execution have in fact prompted the suggestion that intentional actions could be generative, in the same way in which linguistic expressions are. Indeed, authors like Jackendoff went even further, arguing in favor of a neurobiological identity between language processing and action processing. Without necessarily endorsing such strong claims, Mazzone defends this view against a criticism raised by Sperber and Wilson. They argued that ordinary intentions are much more constrained than language meanings, thus lacking precisely the generative nature of the latter—as they put it, "we can say so much more than we can do". Mazzone criticizes this argument on two counts: it seems to underestimate the complexity of non-communicative intentions, and it also presupposes that communicative intentions expressed by linguistic utterances have the same semantic structure as those utterances.

The last two papers have a broader scope. Bruno Verbeek discusses the normativity of intentions, and tries to spell out why failing to do as intended is not only a causal failure, but also a normative one. He considers three alternative types of view. On the first, the fact that you do not execute your intention to do A is 'blameworthy' only if the balance of reasons pointed to A-ing, where intending to do A did not add to the reasons for A-ing. On the second view, the fact that you do not execute your intention to do A is blameworthy because you violate a requirement of rationality. Verbeek points out that both views deny that intending to do A creates a reason to A. Then he moves on to a third alternative, the one he endorses, according to which an agent does create reasons to A by intending to A. He defends his proposal against the well-known bootstrapping objection that intending to A cannot magically create a valid reason to A (at least, the process is in need of explanation). He uses an analogy with political authority to suggest that the normative force of an intention derives from the authority (epistemic or coordinative) of the agent who forms it.

In the final contribution, Cristiano Castelfranchi proposes an analysis of intentions and intentional action within the more general framework of goal theory. He first presents a systematic analysis of the various steps of goal

processing and intention creation, as the final outcome of goal-driven action generation, to argue that intention theory is to be grounded in goal theory, insofar as intentions require means-end reasoning, planning, conflict resolution, and coherence. Castelfranchi emphasizes how deeply intention formation and intentional action execution are regulated by specific sets of beliefs: predictions, evaluations, calculation of costs, responsibility beliefs, competence, etc. He also stresses that the origin of an intention is not necessarily a desire (which is just a kind of goal), and that intention is best understood as a two-layered goal structure: the intended action to be executed, and the intended outcome motivating that action. Finally, he discusses the relations and differences between intentions "in agenda" (future directed intentions; prior intentions) and intentions under execution (intentions in action), and he speculates on how the will is more than the intentions behind our actions.

Taken together, these papers provide up-to-date examples of the many ways in which philosophical analysis and empirical findings can complement each other in the study of intentions, as well as in other areas of philosophical and scientific interest. They also demonstrate how contemporary scholars from various disciplinary backgrounds (philosophy, neuroscience, psychology, linguistics, economics) can engage in productive interdisciplinary ventures, conducted with methodological and conceptual rigor. On the one hand, philosophers are becoming increasingly knowledgeable on experimental findings and methods, and they are very adept at interpreting them. On the other hand, cognitive scientists increasingly strive for conceptual clarity and use experimental results as a stepping stone to generate broad theoretical models. Thanks to these mutual efforts, the current experimental investigation of psychological phenomena is no longer systematically hindered by conceptual confusion, *pace* Wittgenstein.