Rational Agency

Eric Marcus
Auburn University
(forthcoming in *The Routledge Handbook for the Philosophy of Agency*)

According to the Aristotelean definition, the human being is *the* rational animal. Many would for this reason reject it, contending that the rational patterns exhibited by human behavior and cognition are present, even if to a lesser degree, in non-humans as well. One way of defending the definition is to argue that the practical and theoretical capacities of humans constitute a distinctive form of agency: rational agency. Unlike many of the topics in this handbook, this is a relatively unexplored area of research. There is no extensive literature to survey and no warring factions whose battle-lines might be usefully charted. Since many philosophers simply have no idea what a rational form of agency would be, this entry will focus on the very idea.

In a suitably broad sense of *activity*, one might say that activity marks the distinction between the animate and the inanimate. Living things are active insofar as they grow and reproduce. Animals are *self*-movers, and as such are active in a more demanding sense. Self-movement can be understood in terms of an animal's acting on the basis of perception in pursuit of its aims. The network of capacities underlying this activity is constitutive of animal agency. But animals are passive in a way in which humans are not. Whereas we are capable of stepping-back and reflecting on the goodness of our aims and making a judgment about what to do on the basis of this reflection, non-rational animals are driven entirely by their aims. (See Korsgaard 2009.) Similarly, a non-rational animal is not capable of believing on the basis of consideration of what *to believe*, but rather simply believes what it, so to speak, finds

itself believing. This is to gesture in the direction of the idea of a distinctively *rational* agency.

We can begin to flesh the idea out by considering the nature of the capacities that manifest themselves in episodes of rational reflection. Such episodes involve what one might call self-conscious engagement with normative questions, questions paradigmatically of whether to do x or believe p. If a poker player is deciding whether or not to call a large, surprising bet, she is asking herself questions about, e.g., what her opponent's behavior on prior rounds of betting indicates, the significance (and ingenuousness) of his body language, what she knows about his general tendencies as a player, what impact her possible actions will have on her 'table image', whether what might ordinarily be an overly cautious fold would be prudent given how close she is to 'making the money', etc. These reflections require the possession of the concepts that figure in the specification both of the actions being considered and the facts on which the relative wisdom of the various choices depend.

Someone who engages in such reflection does not merely manifest the possession of concepts in the sense in which a dog, who reacts excitedly upon seeing his master reach for the leash, manifests the concepts of 'leash' and 'walk'. A rational subject's manner of possession of these concepts is linked to an understanding of the contributions they make to truth-conditions, to some awareness of the evidential significance of their exemplification, and to a capacity to grasp and evaluate the corresponding propositions. These are abilities that it is not unreasonable to think are limited to language users. We are, after all, unwilling to credit someone with thoughts of the sort that figure in the above reflections if the thinker displays no mastery of the linguistic forms that are used to express them.

Acts of reflection also require the possession of the concepts that articulate the form of the questions themselves. Our gambler must understand the bearing of the considerations she brings to consciousness on questions of whether to *believe* something, e.g., that her opponent is bluffing, and so whether to *do* something, e.g., call the bet. She must thus possess the concepts of *belief* and of *action*. Furthermore, she must understand the very idea of considerations showing a proposition to be one that should be believed, i.e., to be true, and the very idea of considerations showing an action should be performed, i.e. to be good. For this understanding is an element in the relevant forms of rational responsiveness. The subject must thus have some (very general understanding) of the framework of theoretical and practical justification. We might summarize these points by saying that rational reflection depends upon the sort of grasp of concepts that makes it possible to explicitly consider normative questions. It is, in this sense, a self-conscious activity: a rational deliberator understands what she's doing as aiming to make up her mind about what to believe or do.

Suppose this much is granted. Suppose that it is also granted that non-human animals lack the cognitive wherewithal to engage in rational reflection, so understood. Finally, suppose it is allowed that such reflection substantially affects the lives of the deliberators. It would *still* not follow that humans possess anything that deserves to be called a distinctive form of agency. For it might nonetheless be contended that both human and non-human animals act on the basis of prior thought in the same sense, that the difference between them is only in how conceptually sophisticated those thoughts are. This, some will argue, is just a difference of degree. And so what we really have is just ordinary animal agency, which varies in the contents of the associated mental states according to the relevant creature's cognitive abilities. The nature of human thought

and action and the way the former affects the latter is not, it will be concluded, tied to any of the specifically rational capacities that we are now supposing belong exclusively to human beings. Thus we still have no case for or even a clear conception of rational agency.

Our imagined skeptic of rational agency is right about this much: for the possession of a network of capacities to amount to a form of *agency*, their exercise would have to be the central element in a distinct kind of change, distinct in the way that, for example, growth and self-movement are distinct kinds of change. One might express this point by saying that for rationality to amount to a kind of agency is for there to be a distinctive form of causation: rational causation. (If 'causation' seems too metaphysically loaded, then read: "distinctive form of causal-explanation: rational causal-explanation".) But what does this mean?

Deliberation culminates, in ideal cases, in the making up of one's mind: to do x, or to believe that p. These conclusions are based on the reasons for action and belief that proved decisive. We often ask after these bases: "Why did you call the bet?". "Because I thought he was bluffing." "Why did you think he was bluffing?". "Because he always smiles like that when he's bluffing." Here the gambler gives rational explanations of her action and belief. It is characteristic of such explanations that what's offered as explanation is at the same time justification. It is an explanation that consists of a justification. But it is not just *a* justification, it is *her* justification: what, at the time, the deliberator took to justify the relevant belief or action. It was, according to the explanation given, in virtue of her taking the stated reason to justify that the action was performed or the belief adopted.

It has vexed scores of philosophers working in various subfields to say more precisely what it is for the justification to be *hers*. Donald Davidson famously points out that (in the practical case) it is not just a matter of the action being caused by the agent's desiring an end and believing that the relevant action will bring the end about. His infamous mountain climber does not let go of the rope, killing his companion, *in order to save himself*; rather, the recognition that he could save himself by doing so stuns him, thereby causing him to let go. (Davidson 1963) This ('deviant') causation circumvents rather than serves his agency, even as it leads to the realization of his aim.

We have, on the one hand, paradigmatic exercises of rational powers: the making up of one's mind to do x or to believe p on the basis of considerations. We have, on the other hand, certain causal phenomena: someone performs an action or adopts a belief for reasons. There is, apart from any interest in our topic here, a general sense that we are still in the dark as to how to understand the connection between the former and the latter. There would be rational agency if the solution to this mystery went as follows: facts of the former sort *constitute* facts of the latter sort. That is, if the causal connection between a reason and the action or belief that rests on it were nothing beyond the exercise of rational powers—specifically the power to make up one's mind about what to do or believe on the basis of considerations—then there would be a manner of bringing things about that belonged exclusively to rational creatures. Rational agency, then, is (or would be) the power to decide normative questions in a manner so as to constitute facts about the causes (or, if you prefer, causal explanations) of certain events and states. For a subject to accuse the butler of doing it because the butler lacked an alibi or to believe that the butler did it because the butler lacked an alibi is, on this view, nothing over and above the subject's viewing the action of accusing the

butler as *to be done* or the proposition that the butler did it as *to be believed* in light of his lacking an alibi.

According to this proposal, in saying that someone is x-ing or believes that p for the reason R, we attribute to her a normative judgment that x-ing is to be done or p is to be believed in light of R. This is closely connected to the 'Guise of the Good' thesis, according to which acting involves a normative judgment that the action is good. Here the idea is that in acting for a reason one performs the action under the guise of R's contributing to its goodness. A defense of this approach will thus need to reply to those who contend that someone might act for a reason they knew did not establish the goodness of the relevant action—as, it might be contended, a certain sort of akratic does—or for a reason they took to establish the badness of an action—as, it might be contended, an aspiring super-villain does. (See the essays in Tenenbaum 2010 for a recent discussion.)

The thesis that in inferring p from q, one draws the conclusion in the light of the support q provides p is referred to by Paul Boghossian as the Taking Condition. (Boghossian 2014) Here too defense will be required: against the critics of the Taking Condition, who argue variously that there is no way of spelling it out without regress, that it requires too much conceptual sophistication, that it makes false claims about the phenomenology of inference, or that it ignores the possibility of the doxastic equivalent of akrasia. (See, e.g., McHugh and Way 2006.)

The most controversial aspect of the thesis of rational agency is the central causal claim: that certain practical and theoretical normative judgments *constitute* the causal-explanatory nexus between the reason and the relevant act or state. This idea runs deeply against the grain of conventional thinking about causation, according to which

the obtaining of a causal connection is the wrong sort of thing to be constitutively dependent on a subject's representing the world a certain way. It is, of course, uncontroversial that a subject's representation can influence and be influenced by the world. But rational agency, as I suggest it must be understood, involves the idea that a subject's representing R as conferring to-be-done-ness on an action or to-be-believed-ness on a proposition constitutes the obtaining of a causal connection between R and the relevant action or belief. The resultant conception of causation must be defended. (See Marcus 2012).

To think of people as possessing distinctively rational agency is to see them as authors of a distinctive kind of change, one that is tied essentially to their rational sensitivity to reasons. The agent, in concluding that x is to be done on the basis of R thereby *is* x-ing on the basis or R. The subject, in concluding that p is to be believed on the basis of R thereby believes p on the basis R. It follows that there is no gap between R and the action or belief—no 'Reasons Gap', as we can call it. It's not that, say, one judges that one should put out the trash and this causes, via non-rational mechanisms, certain bodily movements that accomplish one's aim. One's judgment that it is to be done constitutes the fact that one is taking out the trash. And so it is precisely this judgment that then makes it possible to perform further actions *because one is taking out the trash*. (See Thompson 2004.) The conclusion of practical reasoning, as Aristotle held, *is* an action. It is the action of x-ing itself that is constituted by the subject's judgment that x is to be done. Similarly, it is not that one judges that one should believe that it's Tuesday, and that this then contributes to the formation (in an optimal case) of the disposition in which believing that it's Tuesday consists. One's judgment that it is to be believed constitutes the fact that one believes it's Tuesday. And so it is precisely this

judgment that makes it possible to then infer that it's not a weekend because it's Tuesday. The conclusion of theoretical reasoning *is* belief, and thus it is belief that p itself that is constituted by the subject's judgment that p is to be believed. Human action and belief are, according to the thesis of rational agency, themselves manifestations of the very capacities exercised in explicit acts of critical reflection.

The thesis of rational agency also eliminates what we can call "The Knowledge Gap". To accept the gap is to think that what one knows simply in x-ing for the reason R or knows simply in believing that p for the reason R falls short of the fact that one is x-ing because R or believes that p because R. To reject this gap is thus to hold that one who is x-ing on the basis of R or who believes that p on the basis of R knows, simply *in* doing so, that she is x-ing because R or believes that p because R. Anscombe's *Intention* begins from the assumption that there is no Knowledge Gap. She argues that someone who performs an action can 'give application' to the rational 'why?', and not on the basis of observation or evidence. (Anscombe 2000.) The agent, on her view, knows *non-empirically* what she's doing and why (in the relevant sense) she's doing it. This claim can be made with equal plausibility about belief and our reasons for belief. Insofar as action and belief are constituted by judgment, we can begin to make sense of this knowledge. Since they are *constituted by* normative judgments, rational creatures have the kind of epistemic access to action and belief that they have to their own judgments. And if there is nothing more to a causal connection between someone's x-ing for the reason that R or believing that p for the reason that R than her viewing R as conferring to-be-done-ness on x-ing or to-be-believed-ness on p, then this non-empirical knowledge of our reasons would be intelligible. The subject can speak with special authority about the question of why the action is performed or the proposition believed

because the relevant facts are constituted by how she herself takes matters to be, specifically: her taking the action as to be done or the proposition as to be believed in light of the justification provided by the relevant reasons.

It should now be evident that there are many challenges facing the defender of rational agency. As mentioned above, it will be argued that people act and believe against their normative judgments, that such judgments are neither necessary nor sufficient for belief and action. The Aristotelean doctrine that action is the conclusion of practical reasoning can seem especially problematic, given that people may never get around to doing what they sincerely and non-akratically judge is to be done, and that people may make such judgments where the relevant action simply cannot ever be performed. Others will hear in the doctrine the implausible suggestion that what we do takes place inside the mind. But, as I have emphasized, the idea is rather that an action must be understood as itself a manifestation of (as opposed merely to an effect of) human practical rationality, as the agent's answer to the question "What should I do?". Actions are, on this view, elements of the space of reasons, and not simply via the proxy of causally efficacious psychological states.

Though Anscombe and those who follow her take the rejection of the Knowledge Gap as a datum to be explained, others would reject it outright. People who are self-deceived might be described as not knowing why they believe and act as they do. Implicit bias might be interpreted as a matter of possessing beliefs of which one had no awareness. And an agent who acts, e.g., selfishly, might prefer to believe a more favorable account of her own action and this in turn might, through selective attention and motivated reasoning, lead her to adopt one. To claim as an advantage a superior position from which to explain our non-empirical knowledge of our actions, beliefs and

the reasons on which they are based, an advocate of rational agency would have to deal with the apparent threat to the datum posed by these sorts of cases.

If acting and believing for reasons require—or, more precisely, if the correct application of the corresponding linguistic forms *to people* require—that the agents and believers possess sophisticated conceptual understanding, then it follows that non-rational beings do not (in the same sense) believe or act for reasons. Some account will then be required of what we are talking about when we describe a dog, e.g., as running down the stairs because its master called for him. Such descriptions impute to animals actions performed on the basis of thought. It is unlikely that any view of the sort I describe in this entry will find a widespread following until a plausible account of non-rational thought and action can be formulated. This remains the most significant challenge for advocates of rational agency.

Although a detailed account has yet to be given, the form that a satisfying response will take is clear. John McDowell argues that although rational and non-rational creatures are perceptually sensitive to their environment, the perception of rational creatures necessarily draws into operation conceptual capacities that the non-rational lack. (McDowell 1994.) Continuing along this path we might say that whereas thought quite generally puts the thinker in cognitive contact with the world, mediating between perception and action, when a species or individual acquires the suite of cognitive abilities that constitute rationality, their perception, thought and action are not merely supplemented with additional contents, but transformed into qualitatively different capacities. When a rational creature sees, thinks and acts, she exercises distinctively rational conceptual powers (in the manner sketched above). A successful elaboration of this idea must exhibit the commonalities that make rational and non-

rational agency instances of a common genus and the differences that make them distinct species of that genus. (Cf., Boyle forthcoming.)

Above, I said that episodes of rational deliberation involve self-conscious engagement with normative questions. Such episodes are marked by the requirement of conceptual sophistication and culminate, in ideal cases, in causation-constituting normative judgments that utilize this sophistication and exhibit the just-discussed rational self-consciousness. But there is no requirement that such judgments be preceded by deliberation. In fact, it is surely the exception. Work needs to be done to understand the nature of these judgments, and what underlies the mutual interdependence of their salient characteristics—that they require conceptual sophistication, that they are made as if in answer to questions about what to believe or do, that they put the rational being in a position to speak authoritatively about the relevant causal matters. But this further understanding will not contradict the self-evident truth that the sorts of judgment in which deliberation culminate are sometimes made without deliberation. It's not as if someone whose deliberation leads him to say that p must be true in light of R is expressing something different from another who makes the same claim without deliberation.

Note that it does not follow from the fact that much of what we believe and do for reasons is not the result of episodes of deliberation that someone who is incapable of engaging in conscious episodes of deliberation could make the relevant judgments. It is only to say that those judgments, with the interlocking characteristics that we introduced by way of considering rational deliberation, can occur even without such deliberation. In fact, it is highly implausible that a creature that is incapable of explicitly

taking up normative questions could nonetheless possess the ability to make judgments that are the taking of stands on them.

Finally, it is worth emphasizing that this sketch of the very idea of rational agency wards off two common misunderstandings of the thesis that humans are a distinctive kind of agent in virtue of their rationality. First, it is not equivalent to the absurd thesis that humans are unerring optimizers and Spock-like cogitators. And not because these are ideals of which we fall short. Rather, these archetypes simply do not personify the rationality that figures in the thesis under discussion. *Rational*, in the relevant sense, does not contrast with *irrational*, but with *non-rational*. Second, the thesis is not that, over and above the exercise of non-rational cognitive powers, we are also capable of exercising rational cognitive powers. This flawed conception would be that while we go about our ordinary business, we think and act in the manner of the non-rational. Then, in occasional episodes of critical reflection, we exercise a capacity that is exclusively our own. These exercises can then impact what we believe and do, where these thoughts and actions are still understood as of the sorts of states and events that figure in the lives of animals more generally. The thesis is rather that the thoughts and actions of human beings quite generally are themselves manifestations of the very capacities that occasionally also manifest themselves in episodes of explicit critical reflection.

Related Topics:

Agency and Causation, Mental Agency, Agency and Practical Knowledge, Agency and the First Person, Agency and Autonomy

Recommended Readings

Boyle, M. and Lavin, D. (2010). 'Goodness and Desire' in Tenenbaum 2010, p 161-202.

The essay argues that the specific form that goal-directedness takes in rational creatures is the capacity to act under the guise of the good, i.e., to act in light of her answer to the question 'What should I do?'.

Boyle, M. (2011). '"Making up Your Mind" and the Activity of Reason'. *Philosophers' Imprint* 11.

Rational creatures can make up their minds about what to believe on the basis of reasons in favor of believing. According to this essay, a belief just is the actualization of this power.

Marcus, E. (2012). *Rational Causation*. Cambridge, MA: Harvard University Press.

An articulation of a conception of rational agency along the lines presented in this entry and of the challenges it poses to the naturalistic orthodoxy in the philosophy of mind.

Moran, R. (2001). *Authority and Estrangement*. Princeton, NJ: Princeton University Press.

An exploration of the connection between the first-person perspective and the phenomenon of normative self-consciousness: our (fallible) ability to know what we believe or want (and why) in virtue of it reflecting our judgments about what is to be believed or to be wanted (and why).

Rödl, S. (2007). *Self-Consciousness*. Cambridge, MA: Harvard University Press.

It is argued that the determination of one's belief as a matter of what to believe and one's action as a matter of what to do is at the same time an exercise of the ability to think of oneself as a subject—as 'I'.

References

Anscombe, G.E.M. (2000). *Intention*. Cambridge, MA: Harvard University Press.

Boghossian, P. (2014). 'What is inference?', *Philosophical Studies* 169 (1),p 1-18.

Boyle, M. (Forthcoming.) Essentially Rational Animals, in *Rethinking epistemology*, ABEL, G. and CONANT, J., eds., Berlin, Germany: Walter de Gruyter.

Davidson, D. (1963). 'Actions, reasons, and causes', *Journal of Philosophy* 60 (23), p 685-700

Korsgaard, C. (2009). 'The Activity of Reason', *Proceedings and Addresses of the American Philosophical Association* 83 (2):23 - 43.

McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.

McHugh, C. and Way, J. (2016). Against the Taking Condition. *Philosophical Issues* 26 (1):314-331.

Tenenbaum, S. (ed.) (2010). *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press, p 161-202

Thompson, M. (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press.