We have many pre-theoretical ideas about freedom or consciousness or knowledge. But we probably have few, if any, about intentional agency as such. The intentional agents that we consider are usually also persons or moral agents. And it is the features that qualify them for these categories that tend to catch our eye. And within the philosophy of agency there is more energy devoted to understanding intentional actions than to understanding intentional agency.

But what intentional agents are deserves our attention. Within the philosophy of agency controversies about intentional action sometimes turn on claims about what role the agent must play in intentional action and whether intentional agency can be captured in, say, event-causal terms (e.g. Mele 2003, Chapter 10). Outside of the philosophy of agency, questions about intentional agents arise in many contexts. Advances in artificial intelligence have given rise to machines who produce complex behaviour. But are they intentional agents? And for that matter, if we, adult humans, our children, and our pets are intentional agents, what qualifies us for this? Without an understanding of the concept of intentional agent and the category of things it delineates, we are at a loss in addressing important questions concerning our similarities to, and differences from, co-inhabitants of our social world.

1. Preliminary Profile: purposefulness, intentional states, explanation

When we think of a human agent building a house or a cat stalking prey or a machine assembling another machine, the behavioural outputs of these things are, we might ordinarily say, purposeful. I will assume that an intentional agent
is something that has capacities for producing purposeful behaviour. What is purposeful behaviour? It is not promising to equate purposeful behaviour with an agent *having a purpose in mind in directing that behaviour*. Someone who scratches out the eyes in a photograph of a rival or who whistles a tune while they work seems to act purposefully without having a purpose in mind in acting thus. (Hursthouse, 1991; Mele 1988) Paradigm cases of purposeful behaviour have structure – purposeful behaviour progresses, not randomly or haphazardly, but along a trajectory, such that some eventualities would count as disruptions of the structure and others would count as proper parts of it. In addition, some parts can be explained in terms of their relationship to other parts. I will assume the following necessary condition on intentional agency:

**Purposefulness:** If S is an intentional agent, then S has capacities that allow her to produce purposeful behaviour.

I will rely on two further assumptions. First, intentional agents are, as the name suggests, beings with intentional states. These are states that represent, or are about, or directed upon, things and states of affairs beyond themselves. Common examples are belief and desire. Second, intentional states help to define this kind of agent by playing a role in generating and guiding their purposeful behaviours.

**Intentional:** If S is an intentional agent, intentional states of S play a role in generating and guiding S’s purposeful behaviour.

Given this role in generation and guidance, a distinctive kind of explanation of the behaviour of an intentional agent becomes possible. This is an explanation of the behaviour that draws on the guiding intentional states:

**Explicability of the Purposeful Behaviour of an Intentional Agent (EPIA):**
If B is a purposeful behaviour of an intentional agent S, then S’s intentional states play a role in explaining B.

Purposefulness, Intentional, and EPIA offer a preliminary sketch of intentional agents. They are a kind of causal unit, one that produces a certain kind of behaviour, and this behaviour is explicable in a distinctive way.

2. **Lower limits on the category**
Bees work together to build hives, make honey, and feed larvae. Given such complex behaviours, they seem to satisfy Purposefulness. Bees’ behaviour seems guided by informational states that, let’s assume, vary reliably with the facts – bees typically register the fact that nectar is nearby, say, only when there really is nectar nearby. (Goldman 1979) Some philosophers think that these informational states qualify as intentional states. (Dretske 1981, 1986) Although questions about the nature of intentional states are relevant and interesting, they go beyond this discussion, so let’s grant this. We can also suppose that the bees’ intentional states concerning nectar play an important role in explaining their behaviours, such as their flight to gather nectar—bees also seem to satisfy EPIA, at least as it is loosely stated above.

We might nevertheless be sceptical that bees are intentional agents. Is this scepticism an objectionable kind of intellectualism about intentional agency? Enactivists argue that creatures even simpler than bees are agents. (Hutto and Myin 2013) Are intentional agents simply that subset of agents who have intentional states guiding their behaviour?

It is not clear that this wide notion of intentional agent would be theoretically useful. Such a category would include insects and higher animals, and it would also include creatures as sophisticated as statistically normal adult humans, who are capable of planning for the future, valuing and desiring things, acquiring skills, and evaluating a wide variety of practical options – courses of action that are open to them to pursue – as morally permissible or not, as prudent or not, and so on. Given this diversity, we would not be able to make too many philosophically interesting generalizations about the members of such a group, such as general claims concerning their distinctive characteristics, their capacities, their moral entitlements, what kinds of action they are capable of, and so on. We would have to look to a potentially high number of sub-categories of intentional agency to frame and answer our questions.

One may worry that there is really no useful category of intentional agent. After all, it does not seem to be a category that significantly shapes our everyday
thought and talk, nor does it loom large in scientific, or legal contexts. But I think that it is relied upon in philosophical contexts. Many philosophers are interested in what makes a rational practical agent autonomous, free, or morally responsible. They presuppose that there are intentional agents who nevertheless fall short of being fully autonomous, metaphysically free, or morally responsible. Human children, sophisticated nonhuman animals, and adult humans acting under duress are some examples. Having a better understanding of this basic general category of agent that we seem to rely on in philosophical contexts would be useful. In what follows, I consider whether there are grounds for excluding bees and other creatures from a category that has real boundaries.

Let's suppose that as long as a bee senses the proximity of nectar, abundant evidence of an insuperable barrier between her and it will not stop her from trying to fly towards the nectar. This behavioural flexibility is notoriously, and somewhat poignantly, reported to be missing from the Sphex wasp. (Dennett, 1984; Wooldridge 1963) This lack of flexibility may be one reason to doubt that the bee is an intentional agent. But human intentional agents may engage in very repetitive—so-called compulsive—behaviour that is not flexible, and yet, we may think that they qualify as intentional agents even when they engage in such behaviour. Given this, it is more promising to think that intentional agents must have capacities for flexible behaviour even if they don't always exercise them:

**Flexibility:** If S is an intentional agent, S has capacities that allow her to exhibit a degree of behavioural flexibility when faced with obstacles to her purposeful behaviour, and this behavioural flexibility helps her to successfully complete her purposeful behaviour.

What capacities must be involved in Flexibility? For starters, the bee would have to have an increased capacity for information-bearing intentional states, and she would also have to have capacities for updating existing information-bearing intentional states in light of new information. This requires ordering the contents of such states so that rational revision of existing informational states becomes possible. If this is correct, we should also accept this further condition on intentional agency:
**Rational revision:** If S is an intentional agent, then S is able to update informational states in light of new information where updating involves rational revision of informational states.

In being prompted to fly towards the nectar, I am imagining that the bee does not have a desire to fly there. Rather, the intentional state registering the presence of nectar triggers this flying behaviour directly. In imagining her in this way, she lacks a kind of intentional state that other agents have—a desire or desire-like state. Roughly, a desire is a state whose primary function is not to accurately represent how things are, but to represent a way that things could be, and to motivate behaviour that will bring this state of affairs into existence. (Schueler 1995) If the bee had a desire-like state whose motivational force were suppressed when there is information to the effect that the desired state of affairs can’t be attained in the current conditions, she could be released from fruitless repetitive behaviour, thereby having an increased capacity for flexible behaviour. Let’s add this to the capacities that are required for Flexibility:

**Desire:** If S is an intentional agent, then S has desires or desire-like states that motivate behaviour in conjunction with informational states.

In addition to increased behavioural flexibility in responding to changing states of affairs, a desire or desire-like state gives the creature greater causal independence from her immediate environment. If we assume that the state does not originate in the immediate environment, it allows the creature to have a source of behaviour that stems from within her. This resonates with the idea that an agent is some kind of causal unit that enjoys a degree of causal independence from its surroundings.

Nevertheless, if the desire-like state is merely given by a creature’s biological nature, or by her designer, and if the motivational force of the state is dampened or extinguished by informational states that are caused by the environment, then the causal independence that she has is quite slight. In fact, she seems to be a “site” of causal processes rather than a source of them. This suggests that an intentional agent should also have some causal independence
from her desire-like states. She should have, at a minimum, a capacity for suppressing desire-like states that is not determined by inputs from her immediate environment. This suggests the following further condition on intentional agency:

**Independence:** If S is an intentional agent, S has a certain degree of causal independence from designer-given or biologically-given desire-like intentional states in the production of her purposeful behaviours.

We have considered what conditions might allow us to make sense of a general category of agent, who falls short of autonomous, free, morally responsible agency, but is nevertheless more complex than a bee. Creatures that lack the capacities that undergird flexible behaviour, such as information-updating that is governed by rational rules, who lack desires or desire-like states, who lack causal independence from their environment, and from their own externally-determined motivational states, may seem to be more like automatons—beings lacking independence from the past or their make-up or their environments—rather than intentional agents.

**3. The agent it/her/himself**

Suppose that a complex computer has the two kinds of intentional state discussed above, and the operation of these states is not wholly determined either by the immediate environment or by a designer-given programme. Suppose that the computer is such that in informational state, IN, there is an 80% chance that a designer-given desire-like state is suppressed and a 20% chance that it will not be suppressed but will motivate and guide a bit of purposeful behaviour. Because of the probabilistic relationship among the computer's states, whether or not the motivational state is suppressed is not directly determined by the designer. Suppose also that the computer produces purposeful behaviours that satisfy its desire-like states and that these behaviours are explained in terms of these desire-like states and its other intentional states. Would this computer qualify as an intentional agent?
Does the computer itself produce these behaviours? Although the computer may be composed of complex causal chains that enjoy considerable independence from its immediate environment and its designer, the computer is not clearly something that itself exercises control. To come at this idea from a different direction, let’s suppose that a paradigmatic product of intentional agency is an intentional action involving bodily movement. But intentional actions are not things that are produced by parts of agents, they are performed by, and attributable to, the agent herself. Intentional actions are “personal” rather than “sub-personal” phenomena, they are performed by the agent rather than caused by an event in the agent. To come back to the computer, does it fulfil the following condition?

**Source:** If S is an intentional agent, then S has capacities that allow her to exercise control over her purposeful behaviour.

Event-causalists about intentional action characterize the agent’s exercising control in intentional action in terms of certain kinds of causal pathway obtaining between mental events, such as desire and belief, or intention (see Bratman 1999; Mele 1992; Sinhababu 2017 for discussions of the psychological states involved), and behaviours that are also events (e.g. Bishop 1989; Enç 2003; Mele 2003; Smith 2010; Shepherd 2014). This is an appealing view, because it does not involve an irreducible or supernatural entity in the role of agent. But there has been a debate about whether event-causal views suffer from a so-called “disappearing agent” problem. (Hornsby 2004, 2010; Mele 2003, chapter 10; Nagel 1986; Schlosser 2011). The event-causal view, it is argued, yields a picture in which one kind of event (mental) causes another (bodily behaviour), but the latter is not an agent’s performing an intentional action. In attempting to reduce the operations of agency to causal relations between mental and bodily events, the agent has disappeared altogether. A related complaint is that typical event-causal views that characterize actions as caused by intentions or desire-belief pairs can only account for actions from which the agent is alienated, rather than actions “par excellence”, roughly, actions with which the agent identifies. (Velleman 1992; Mele 2003, chapter 10)
Two issues are intertwined here. One is conceptual in nature: what is required for the agent herself to perform an intentional action – must she be a controller of her action, or the source of her action, or must she identify with the action, or something else? And what is it to be any one of these things? The second issue concerns what it would take, ontologically speaking, for a creature to be the controller, or source, or to identify with an intentional action. Would the agent have to be a substance who gives rise to, and intervenes in, causal processes, but whose operations are irreducible to causal relations among events? Or can the characteristic activities of the agent be characterized in event-causal terms?

Event-causalists make a number of responses. They stress that the mental events involved in the production of intentional action are agent-involving psychological states: they are thoughts of the agent. Our talk of the computer’s “belief-like” or “desire-like” states, they might say, mislead us—a richer account of the psychology of intentional agency than we have considered so far will assuage at least some of the worry about disappearing agents. And we also need to characterize the ways in which such psychological states play a role, not just in causing movements, but also in guiding these movements as they unfold over time. (e.g. Bishop 1989) It is also argued that the disappearing agent worries rest on muddy intuitions that risk exaggerating the role of the agent, of ignoring continuities in agency between human and non-human animals, and misunderstanding the reductive naturalistic aims of the event-causalists. (Mele 2003, chapter 10; Schlosser 2011) For worries that event-causalism can only capture alienated actions, it is argued that as long as there is a certain kind of psychological complexity in the production of actions, an event-causalist can capture the role of the agent in the production of intentional actions “par excellence”. (Velleman 1992; Mele 2003 chapter 10)

Those who reject event-causal responses opt for agent-causal views (e.g. Alvarez and Hyman 1998; Brent 2017; Mayr 2011), characterizations of agency in terms of two-way powers (e.g. Alvarez 2013), volitionist views (e.g. O'Shaughnessy 1986/2008; Pietroski 2000), and recent work on the ontology of
processes also challenges event-causal approaches to intentional agency (e.g. essays in Stout 2018). Agent-causal views capitalize on, and contribute to, a rich ongoing research programme into substance causation, presenting an alternative to the Humean event-causal picture. Volitionist or trying theories offer accounts of willing that are irreducible to the operation of psychological states, such as desire. All of these views carefully articulate a number of worries about the event-causal programme in mind and action. It goes well beyond us to consider the details of the competing views, but as they centrally concern how to capture the difference between being a mere site of causal processes, and being a performer of intentional actions, they are an excellent place to start in trying to capture what it takes for a condition like Source to be satisfied. One final thing should be noted: although I have here talked about intentional actions involving bodily movements, we must also consider what it takes to capture the role of the intentional agent in the production of other behaviours such as mental actions, omissions, refrainings, and lettings happen.

4. Unity in the category
So far, I have motivated some conditions on intentional agency. Much more must be said to develop such conditions, but let’s turn now to the issue of whether the category, even as briefly sketched, promises to have the internal unity that was hoped for—a unity that would allow us to make fruitful generalizations about its members.

A hungry cat sees a pigeon and begins to stalk it. But then she notices that there is a large dog approaching the pigeon. The cat promptly gives up the pursuit and retreats to lie in the sun. She watches and waits, attending to whether the dog moves to a safe distance. Let’s suppose that when the cat notices the pigeon, her mouth begins to water, she feels excitement, she feels motivated to move towards the pigeon, and these thoughts and feelings play a role in causing her to stalk it. When she notices the dog, she finds it threatening, she feels fear, and these thoughts and feelings cause her to stop in her tracks, to crouch down, and retreat.
In breaking off her pursuit of the pigeon, let’s suppose that the cat doesn’t judge, nor does she have the capacity to judge, that the threatening dog would make her pursuit of the pigeon a mistake. She doesn’t, that is, treat the presence of the dog as a reason to retreat. (Kauppinen forthcoming; Schlosser 2012) And let’s suppose that she doesn’t think that it would be bad or a mistake for her to cross the dog’s path. Furthermore, let’s suppose that the cat doesn’t have either the capacity to recognize that there is more than one practical option open to her, or the capacity to deliberate about which of two or more practical options to take, finally making a decision about which one is best. Rather, her finding the pigeon desirable or the dog threatening leads directly to attentional and behavioural changes that constitute her goal-directed behaviour of stalking the pigeon or avoiding the dog.

Contrast the case of the cat with the case of a hungry child who sees a juicy apple. It looks delicious to her, her mouth begins to water, she feels excitement, she feels motivated to get it, and these thoughts and feelings play a causal role in her reaching for it. As she reaches, she notices that her friend frowns at her and begins to utter words of protest. The child hesitates. Let’s suppose that her hesitation is not just a conditioned response to frowns and protests, she has matured beyond that. Rather, she thinks that she shouldn’t take the apple because it is her friend’s, that it would be wrong of her to do so. These thoughts and feelings play a causal role in her bringing her action to a halt. After a little more thought about what to do, she decides to leave the apple to her friend.

When looked at “from the inside” the cat and the child are very different: the cat doesn’t take states of affairs as reasons to act or not to act, the child does, the cat does not engage in reflexive thought, the child does, the cat doesn’t recognize that there is more than one practical option to choose from, whereas the child does. And the cat doesn’t regard herself as potentially criticizable, but the child does. Although they both plausibly fit the conditions on intentional
agency that we have considered so far, their differing capacities for deliberative, normative, and reflexive thought make them very different from one another.

Should we tighten the conditions on intentional agency to exclude the cat, say? For example, should it be required that intentional agents have the following capacities?

**Reasons:** If S is an intentional agent, then S has the capacity to recognize and respond to reasons as such.

**Decision:** If S is an intentional agent, then S has the capacity to decide to take a course of action from among two or more courses of action that seem to her to be open to her.

Alternatively, should we use such conditions only to specify sub-categories of intentional agent? I will next offer some reasons to think that we can have a meaningful general category that includes creatures such as our imagined cat, the child, and also sophisticated adult human agents. Reasons and Decision could be used to delineate sub-species of the category of intentional agent, but the diversity would remain tolerable because of significant similarities across otherwise diverse agents.

In spite of their differences there are striking similarities between the cat and the child: both have first-person perspectives on the world such that there is something that it is like to be them as they navigate that world. (Nagel 1986) The phenomenal aspects of their intentional states seem to play roles in prompting and guiding their actions. Note that when adult human agents engage in sophisticated practical deliberation and action, this process often originates in an occurrence that is common to cat, child, and adult: they are moved by how things in the world strike them—as desirable or undesirable, as disgusting, delicious, ugly, beautiful, and so on. And even though a cat may not have a capacity for valuing things, for planning to do things in the future, or for self-criticism (Hieronymi 2009; O’Brien 2019), our capacities for these may be grounded on our capacity to find things beautiful, delightful, painful, and so on. (Carruthers 2018; Jaworska, 1999)
It is also worth noting that a first-person perspective may play a role in making behaviours intelligible in a distinctive way. It may be that it is because we, statistically normal adult human planning agents know what it is like to feel very hungry and to find something delicious, that we may grasp in a special way why an agent would reach for it quite spontaneously. And it is because we know what it is like to find something threatening or forbidding, that we grasp why an agent would cower and stop in their tracks. Although the cat and the child may be very different from us in their capacities for normative or reflexive thought, their actions may nevertheless be at least partially intelligible to us in a distinctive way precisely because of similarities in the way things seem and feel to all of us, and because of similarities in how we all act in response. Given this, we should consider whether a condition such as the following is also necessary for intentional agency:

**FPP (First-Person Perspective):** If S is an intentional agent, then there is something it is like for S to navigate the world and these phenomenal aspects play a role in generating and guiding her action.

I am tentatively suggesting that the right way to proceed in understanding the category of intentional agent is that, in spite of diversity in capacities for complex thought, all members of the category satisfy FPP. In addition, we should consider whether earlier conditions, such as Intentional and EPIA should be developed in ways that reflect FPP.

In the earlier somewhat curt introductions of belief-like and desire-like intentional states, they were not characterized in terms of how they feel to the agent. For all that was said, an account of them in terms of direction of fit (Anscombe 1957/2000; Frost 2014), or in purely causal-functional terms (Lewis 1972; Putnam 1975)—terms that makes no essential reference to phenomenal aspects—might have been acceptable. Perhaps zombie minds (Chalmers 1998)— minds in which there is nothing it is like to have a desire or belief, or to pursue some goal—and zombie intentional agents are possible. And perhaps there are artificial agents who lack phenomenal aspects of intentional states, but
seem to be intentional agents. Clearly, FPP is no shoo-in. Nevertheless, it is important to consider whether FPP, together with the other conditions, is one of the keys to the internal unity of a philosophically fruitful category of intentional agent. And we should think carefully about whether any such philosophically fruitful grouping can countenance the satisfaction of FPP by some of its members but not by all.

**Recommended Reading**

Alvarez 2013 develops a view of human agency in terms of two-way powers. Baker 2013, especially chapters 2 and 6, offers a fresh and accessible discussion of the first-person perspective in different types of agent. Bratman 1999, especially chapters 2 and 3, develops the planning theory of intention and defends the irreducibility of intentions to desires and beliefs. Hieronymi 2009 develops an account of the will that emphasizes the criticizability of the agent herself for her intentional actions. Mele 2003, especially chapters 2, 3, and 10, discusses a range of issues concerning the defensibility of a naturalistic event-causal view of intentional action and agency.

**Bibliography**


Kauppinen, A. forthcoming. “Rationality as the rule of reason”, Nous.