Croatian Journal of Philosophy
Vol. XXIV, No. 72, 2024
https://doi.org/10.52685/cjp.24.72.3
Received: May 6, 2024
Accepted: July 4, 2024

Intentions and Representations

SHAUN GALLAGHER
University of Memphis, Memphis, USA

Kathy Wilkes's essays on explanations and representations, and especially her interaction with Daniel Dennett, raise questions about whether some notion of representation can explain action intention. Wilkes is not sure whether subpersonal representations are real, but she thinks that the most pragmatic strategy is to take the intentional stance and accept the usefulness of personal level intentions, even if we have to worry that this does not give us a scientific explanation. Wilkes's skepticism about subpersonal representations, and even about the appropriateness of the notion of subpersonal levels of explanation, seems to fit with more recent embodied-enactive approaches to cognition. Considerations about the nature of cognitive mechanisms and animal intelligence prevent her from moving in that direction, however. These insights suggest that Wilkes' analysis continues to be directly relevant to contemporary discussions.

Keywords: Action intention; representation; subpersonal levels of explanation; intentional stance; enactivism.

1. *Introduction*

My aim in this paper is to look at some things that are missing from Kathy Wilkes' essay "Representations and explanations" (1989a) and to see if by bringing those missing items into the discussion it could clarify some of the problems she is considering, and also give us some idea of how Wilkes would fit into some current debates about representation. Some of the missing things are missing simply because Wilkes ignored them; other things were not yet available when she wrote her essay. I'll also make reference to a second essay she published in the same volume, "Explanations—How not to miss the point" (1989b). Both of these essays are part of a volume, *Goals, No-goals and Own Goals. A Debate on Goal-directed and Intentional Behaviour,* based on a set of seminars that took place in Oxford in the 1970s and 1980s, as Monte-

fiore and Noble indicate in their editorial Introduction to the volume. Due to the nature of the volume, she makes reference to other essays in the volume (by Dennett, Montefiore, McFarland, and Noble), and she is explicitly in dialogue with Dennett on some specific points. It's notable, I think, that neither Wilkes nor any other contributor to this volume, which explores issues having to do with intention, makes mention of well-known work by Elizabeth Anscombe (1957) or John Searle (1983) on these issues—even to disagree with them.

2. *Some stage setting*

Wilkes, in her essay "Representations and explanations" (1989a), is concerned with the notion of intention, in the sense of having or forming an action intention, understood as a representation of a goal to be attained. Her question is whether one needs these concepts (intention, representation) to explain behavior. Her answer ultimately is yes with some important qualifications. To work out this answer she explores the notion of explanation and along the way discusses, and endorses, Dennett's intentional stance (without naming it as such).

One issue that we might be tempted to set aside, not only because it seems to be a terminological issue, but because Wilkes herself sets it aside, concerns the distinction between intentionality (with a 't') and intensionality (with an 's'). For her, intentionality with a 't' signifies "the forming or having of intentions, representations of goals to be achieved" (Wilkes 1989a: 159), and this is taken to be a sub-category of intensionality (with an 's'). Wilkes (1989a) equates intensionality (with an 's') with what most would consider Brentano's concept of the aboutness or object-directedness of mental states, which is usually spelled 'intentionality' (with a 't'). Wilkes, then, does not use the standard or orthodox understanding of this terminological distinction.[1] According to Searle, for example, this is something of a mortal sin:

> One of the most pervasive confusions in contemporary philosophy is the mistaken belief that there is some close connection, perhaps even an identity, between intensionality-with-an-s and Intentionality-with-a-t. Nothing could be further from the truth. They are not even remotely similar. Intentionality-with-a-t is that property of the mind (brain) by which it is able to represent other things; intensionality-with-an-s is the failure of certain sentences, statements, etc., to satisfy certain logical tests for extensionality. The only connection between them is that some sentences about Intentionality-with-a-t are intensional-with-an-s […]. (Searle 1983: 24)

Whether we accept Searle's characterization or not, it does seem that in contemporary debates about representation, the distinction holds some significance; in some broad sense it relates to an issue discussed by Wilkes, about whether having a representation with semantic con-

[1] Despite the fact that she cites Chisholm (1957), the likely source for the orthodox view. "Any reader who wants a short and clear description of what 'intensionality' is should consult chapter 11 of [Chisholm 1957)]" (Wilkes 1989a: 182).

tent depends on having linguistic ability. We'll return to that issue. For now I'll adopt the more standard distinctions between action-intention, mental intentionality (aboutness), and language-based semantic intensionality.

At this point, however, Wilkes is more concerned to distinguish action-intention, from intentionality in the broad sense of aboutness. Action-intention is a special form of the more general concept, of course, since action-intention also has the character of being about something or being directed towards something (e.g., a goal).

3. *Levels of explanation*

One issue addressed by Wilkes concerns causal explanation, and whether that would be the right kind of explanation for explaining action intentions, or for explaining behavior using intentions as part of the *explanans*. In this respect, there is also some question about levels of explanation. An action intention seems to be a personal-level phenomenon; but a representation is sometimes considered a sub-personal thing. Consider that neuroscientists keep telling us that they can identify what someone is thinking, or perceiving, or intending to do by looking in that person's brain using e.g., fMRI (e.g., Coles 1989; Cox and Savoy 2003; Frith and Gallagher 2002; Haynes et al. 2007). Chris Frith, for example, thinks that an intention is part of the neural mechanism involved in motor control for intentional action, specifically a representational part of the comparator process which, to keep an action on track checks the match between intention and efferent copy (generating a sense of agency) (Frith 1992). One can find similar ideas in discussions of predictive processing (Friston and Frith 2015; Gallagher and Allen 2018). As I understand her, Wilkes wants to rule out such explanations, i.e., any explanation that would treat intentions as subpersonal representations with causal power, and she wants to limit the notion of intention to the personal level.

With respect to the issue of explanation, Wilkes cites an example from Hilary Putnam (1980): why the square peg won't fit into the round hole cannot be explained by a molecular-level explanation. That would be the wrong sort of explanation. Just as molecular processes do not cause the mismatch between square peg and round hole, neural processes do not cause the intention—"*even if* intentions, or representations, can ultimately be described as 'no more than' sets of processes in amongst nerve cells" (Wilkes 1989a: 169). This would be a statement of composition rather than causality. Neural processes do not cause intentions, and do not causally interact with intentions; even if they constitute intentions (on some kind of identity theory).

I think we get closer to her reasoning by considering her own example: a neural explanation won't explain why Flora flounced out of the party—one needs to explain that behavior in terms of intentions, beliefs, desires, etc. in what Sellars would call the space of reasons (folk

psychology) rather than the space of causes. No one in this collection of essays mentions Sellars either. No need to, since we have Dennett and the intentional stance with its distinction from design and physical stances (Dennett 1989: 237). These stances reference different levels of explanation: the personal- or intentional level versus the subpersonal, distinguished into functional (design) and physical levels. Things get complicated, however, since explanatory levels can be defined in different ways (Wilkes 1989b: 198–199); mereological/constitutional *versus* functional/causal for example.

This kind of complication is discussed in more recent philosophy of science. Phillip Gerrans itemizes a number of level types.

> The notion of levels is ubiquitous [in scientific explanation], but not everyone uses it in the same way. It can refer to ordering relationships between theories…; the objects of theories ordered by size or complexity, e.g., cells are smaller and less complex than the organs they make up; functional analyses, e.g., vision is a higher-level property than edge detection; or levels of mereological containment, e.g., parts are at a lower level than wholes. (Gerrans 2014: 229–230)

Kenneth Schaffner (2020: 384) provides a more exhaustive list of level types: levels of "abstraction, analysis, aggregation, behavior, complexity, function, perspective, organization, generality, and processes—as well as causation and control—as well as description and explanation, and more."

Given this complex multiplicity of levels, James Woodward (2020: 428) expresses an attraction to "levels eliminativism […] the thought that we would be better off avoiding level talk entirely." He nonetheless introduces what he calls the 'interaction level'. He takes the interaction level to include any factor, regardless of size or composition, that has a causal relation to the system that needs to be explained. Such factors are testable by his notion of interventionist causality. This puts neural processes, psychological processes, social processes, etc. all on the same level, so that the explanation doesn't have to reference any other level (defined by different criteria). In this sense, if one's explanation is confined to one level, so defined, then one doesn't have to talk about levels at all.

This approach might have appealed to Wilkes since she had a skeptical view of levels and a quite pluralistic or "tolerant" (1989b: 209) view of causes ("We describe things as causes when they interest us, when they seem important to us, when we can juggle and manipulate them" [1989b: 205]).

> In psychology and neuroscience …we have practically no idea what, and where, the relevant 'levels' between (at one extreme) the macro-states postulated by the behavioural sciences, and (at the other extreme) the individual synaptic connections described by neurophysiology, are. We lack an agreed neuropsychological taxonomy 'in the middle'; and, as noted already, psychology at the 'macro' level still has little consensus about its taxonomy of explananda. The top 'level' is very loosely characterised as yet; and the levels beneath that are still largely matters of mystery. Thus we do not

> know whether 'intentions', or 'goal representations', have a suitably systematic relation to anything on the next level down (whatever that might be). (Wilkes 1989a: 174)

I think this is still the case, but also I think that Wilkes gets tripped up about levels by her examples, which are examples of physical level mereological relations rather than functional-causal relations. That's clear in her appeal to Putnam's example of the square peg; and also her example of the ornament. An intention is like an ornament, but "the way in which atoms and molecules constitute ornaments is not something amenable to scientific investigation; and for that very reason 'ornament' is not an explanandum for physics" (Wilkes 1989a: 171)—at least in regard to its cultural significance. If the complications about levels of explanations lead her to endorse the intentional stance, this is not the only reason. Another reason is her uncertainty about the reality or status of representations. In this regard Wilkes argues that "it is vastly unclear what it means to say that 'there are such things as' intentions, or goal representations. Yet if they are to be worth citing as 'causes' in the explanation of behaviour, then, evidently, they must exist. If they have a role in explanation, but not as causes of behaviour, then the matter is less clear" (Wilkes 1989a: 170).

4. *Are representations real?*

Here, Wilkes engages with Dennett and the idea that representations must be "physically structured objects" that play a causal role in cognition. Quoting Dennett: "... information is represented explicitly in a system if and only if there actually exists in the functionally relevant system a physically structured object ..." (Dennett 1982-3: 216; Wilkes 1989a: 161). Specifically, for Dennett, a representation is a physically structured object plus some kind of interpretation or interpretive mechanism. The two together realize a representation. But, Wilkes is hesitant: "what it means to say that representations or intentions exist—is a highly vexed business."

One interpretation of the Dennettian view is that a representation is a kind of subpersonal entity, explicable from a design stance. This would be distinguished from whatever might (or might not) be on the personal level, grasped via the intentional stance. The scientific explanation is focused on the design level. Wilkes, however, is drawn to the personal level, where an intention is equated with what she calls an 'explicit' representation. For her there is something like a "sliding scale" that descends from an explicit representation (intention) to lower-level operations. Here Anscombe is not mentioned, but an Anscombian analysis seems to be implied:[2]

[2] I have in mind Anscombe's example of the man working a pump. What counts as the action can be described in many ways, including just the physical use of muscles to pump the water. But the circumstances will say what the most appropriate description is. For example, if the water is poisoned and the occupants

> For instance, if one intends some end—killing Lincoln, say—then in a sense one intends the various means to that end; one may be said to intend whatever the guiding intention implies. Thus, perhaps parasitically, Wilkes Booth ('also'?) intended to shoot Lincoln; to fire a gun; to pull the trigger; to crook his right index finger. We are, I think, much less certain about whether these are 'explicit' or not. We might call them 'implicit', in that they are implied by something that is (perhaps) explicit. But we have no clear intuitions about whether, or when, they are 'worth individuating', or what it means to say they 'exist'. The problems of individuating intentions are, unsurprisingly, exactly the same as those of individuating actions […] evidently [we] have a hefty theoretical problem in the specification of 'the' intention, or representation, that guides, governs, explains, modifies, or perhaps causes, purposive behaviour. (Wilkes 1989a: 162)

It's not clear that in this listing of implicit intentions/representations Wilkes (Cathy, not Booth) goes far enough down to get to anything subpersonal. Nothing here resembles a physically structured object of the sort that would count as a subpersonal representation.

Even when she considers what she calls 'tacit representations' they are not necessarily subpersonal, although subpersonal processes are clearly involved—"Tacit representations seem to be dispositions, abilities, know-how: where and what we can do depends upon the way we are, or—sometimes—on what we have learned" (Wilkes 1989a: 163). It's not necessarily the case that individuating actions or intentions remains theoretically problematic once one introduces some of these Anscombian descriptions, but they lead directly to pragmatic considerations about circumstances, embodied degrees of freedom and ecological affordances (Gallagher 2020).

It's still an open question (at least for some) whether one can lower the analysis into the subpersonal scale, to find Dennett's 'physically structured objects.' In the contemporary discussion (unavailable to Wilkes) such objects are called 'structural representations.' Structural representations are described precisely as mechanisms on the subpersonal, neural level. Gualtiero Piccinini (2022: 6), for example, suggests a way to think of such representations as physically structured objects. For him, representational content is just the information contained in the occurrent physical structure of neurons or neuronal networks (which can be understood following Gabor [1946] and Miłkowski [2023] as Shannon information instantiated in the quantifiable independent degrees of freedom of such physical entities):

of the house die, then pumping the water could be a case of murder, depending upon the agent's knowledge and intention. "[A] single action can have many different descriptions […]. Are we to say that the man who (intentionally) moves his arm, operates the pump, replenishes the water supply, poisons the inhabitants, is performing four actions? Or only one? […]" (Anscombe 1957: 11). In short, the only distinct action of his that is in question is this one, A. For moving his arm up and down with his fingers round the pump handle is, in these circumstances, operating the pump; and, in these circumstances, it is replenishing the house water-supply; and, in these circumstances, it is poisoning the household. So there is one action with four descriptions" (Anscombe 1957: 45–46).

> A specific content may be distributed over a relatively large ensemble of neurons. Yet content is relatively localized in the sense that it is carried by a specific vehicle born by a specific bearer (neuron/ensemble/circuit) and not diffused through the whole neurocognitive system, or even a large part thereof. (Piccinini 2022: 6)

If one stays with the concept of Shannon information, then this means that content just is the physical configuration that defines the neuron's function in the neuronal ensemble, the degrees of freedom of a neuronal network, etc. If this does not solve all problems, it nonetheless is a good candidate for Dennett's physically structured object. Questions still remain whether this isn't just a neural structure that covaries with environmental stimuli, and in that sense why we should consider it a system-relative representation rather than an observer-relative interpretation, or deflationary gloss (Egan 2014). Moreover, if the physical pattern is indirectly, yet still physically, coupled to the environment or object correlated with that co-varying pattern, it can be explained in terms of dynamical causality rather than anything resembling good-old-fashioned semantic content. Although Piccinini thinks such a neuronal structure can be decoupled from its target, he also contends that we don't even get this far without the system being embodied, embedded, enactive and affective. That's where the structure comes from. In which case one can ask about the embodied origins of so-called 'non-derived' original content. At the very least, these are questions that we can raise about why we would want to call this a representation in the first place.

A couple of years after the publication of Wilkes' essay, Dennett published his essay "Real patterns" (1991) which suggests an answer to the question, are representations (of this subpersonal type) real? (This is another thing that was not available to Wilkes, although I wonder whether the dialogue here between Dennett and Wilkes didn't directly motivate Dennett's thinking about patterns). Whereas Wilkes has to say she just doesn't know, Dennett would contend that representations are real in a scientific pragmatic sense. That is, they are real enough if science finds them useful components of an explanation—if they serve some pragmatic purpose in empirical explanation. This is not an ontologically heavy conception of reality; it seems to go along with the notion of content that is not of the heavy semantic intensionality (with an s) type, as well as with the notion of intention in the intentional stance, which seemingly does not come along with ontological commitments. Wilkes embraces this latter kind of explanation—the intentional stance—which is just what allows her to remain uncommitted about a subpersonal explanation (Wilkes 1989b: 195):

> [I]t is almost entirely irrelevant what (if any) neurophysiological processes underlie the psychological dispositions or processes which we cite in such explanations—these have no bearing on what interests us […] there may be no systematic correlations between descriptions of intentions and of cerebral processes. Objects picked out by common sense, since they are not

> necessarily (indeed not often) natural kinds, won't usually have any systematic reductive correlations with any microstructural descriptions. (Wilkes 1989a: 168-169)

Even if a representation were radically token-token correlated, perceived object to neural processes, something that could still be simple co-variation, Wilkes doesn't think this is explanatory:

> But more than that: it might put into question the viability of [person-level intentions or] representations as appropriate scientific explananda or explanantia. If representations cannot be explained systematically by states of the brain, what is the scientific justification for postulating them in the first place? One reply, of course, is to say that psychology is 'autonomous.' (Wilkes 1989a: 171)

If person-level intentions can be explained by a physically structured object (a structural neural representation), then person-level intentions would play no part in (or be redundant in) explanation of behavior, and psychology would not be autonomous; if they can't be so explained, then they seem less real. Wilkes is here anticipating and opting for Dennettian pragmatism:

> Nonetheless it is hard to avoid the conclusion that explanations lacking 'intentions,' or 'goal representations' will, by and large, come out as superior to those that possess them […]. To defend such woolly postulates as 'intentions' or 'representations' we'd need to establish that there were instances of behaviour which could not be explained without them; or which could only be explained in a highly unwieldy way without them. This is absolutely crucial, for, if this could be established, then other deficiencies of 'intentionalist' theorising can perhaps be overlooked; I have already argued that it's better to have some explanation than no explanation. (Wilkes 1989a: 176)

5. *Extended mind but not enactivism*

For Wilkes, the concept of intention is clearly the idea of a prior intention, formed in deliberation. Wilkes discusses a representation of a goal that a person forms as the result of some deliberation (going to the bank tomorrow), which action the person then sets aside until tomorrow. There is no mention of intention-in-action when tomorrow comes, something one would find in Searle. Rather, Wilkes compares the representation held in memory to an example that has since become a central example in the idea of the extended mind—setting this decision down in a notebook:

> These [notebook and natural] representations help guide our behaviour. They seem, too, to be phenomena that we want to construe realistically: phenomena needing to be individuated, and which 'really exist' […]. What is special about the 'guiding' of diary entries, or sudden recollections of an earlier decision? Simply that they cannot guide us unless their semantic content is understood. The marks in a diary must have meaning for the user […] (Wilkes 1989a: 181)

Indeed, she seems to anticipate and endorse the idea of extended mind: "Some people do not need diaries, and keep their decisions 'recorded' in

short-term memory—but the difference between written and remembered intentions seems insignificant" (Wilkes 1989a: 181).

This example points to a significant distinction that remains implicit in Wilkes's account—and it may clarify things to make it explicit. Both forms of representations (one the result of biological memory; the other an external representation in a notebook) are *products* of some cognitive doings. They are not representations with original content, or representations found in the mechanistic or physiological processes that may be doing our cognitive work. Representations-as-products may have an influence on such mechanistic or physiological processes in a derived or secondary way, if they loop back into subsequent cognition, or inform our intentions-in-action as we set out to do our action.

Accordingly, there is an important distinction between:

- *Representations as products* of cognitive processes, operating at the personal level—e.g., memory is the representation of a past event—this may involve language—and may be part of a folk psychological explanation.
- *Representations as components of (or processes in) the mechanisms* that explain cognition, operating on a subpersonal level, providing a functional or physical explanation.

For Wilkes, "The earlier deliberation, *resulting* in some intention [representation] or other, seems required by any adequate explanation of the behaviours in question […]" (Wilkes 1989a: 181), at least in a folk psychological explanation.

What's been confusing throughout this discussion is the difference between a representation that is posited as part of the mechanism that produces cognition (these are the putative subpersonal structural representations characterized in causal terms) and a representation as a personal-level intention that is the product of a deliberative cognition. Wilkes is not sure whether either form of representation is real, but thinks, along with Dennett, that the most pragmatic strategy is to take the intentional stance and accept the usefulness of personal level intentions, even if we have to worry that this does not give us a scientific explanation:

> That the explanation of much behaviour can only be given in teleological, intensional idioms, and even idioms that cite 'intentions,' I accept. That behaviour *as classified in ways appropriate to a science of behaviour* can only be handled by reference to […] internal representation [framed in terms of] the neurophysiological ('hardware') nitty-gritty" […] I doubt (with a few qualifications). (Wilkes 1989b: 204–205)

If Wilkes thinks that this is in some agreement with Dennett, Dennett, in a critical response to Wilkes, disagrees:

> I certainly agree that explanations are not all of the same type. I distinguish physical stance explanations, design stance explanations and intentional stance explanations. There are finer distinctions that also seem well-motivated to me, but I don't yet see why we can't use them all in science—and in everyday 'commonsense' explanations. (Dennett 1989: 237)

In regard to current debates about representation, let me note that enactivists who tend to be the strongest anti-representationalists in the embodied cognition camp, would not necessarily object to Wilkes' or Dennett's pragmatic defense of action-intentions, that is as products of deliberative processes, especially if such deliberations involve language. They may add intentions-in-action (P-intentions) and motor intentions (M-intentions, processed at the subpersonal level) to get a fuller story (see Pacherie 2008); but they would object, as Wilkes does, to positing subpersonal representations as part of the explanatory mechanism for any of these intentions.

There are, however, two things that suggest that Wilkes would not be happy moving in the direction of enactivism. The first is her story about the fuel-saving device; the second is her citation of Macphail about animal intelligence.

First, Wilkes repeats a story she learned from Naomi Sheman:

> An advertisement claims that unbeknownst to most drivers (perhaps because the automobile manufacturers are in cahoots with the oil companies), there is a fuel-saving device in all cars, and for a mere $29.95 we will send you what you need to know in order to activate it. When you send in your money what you receive is a set of tips such as: avoid jack rabbit starts, use the highest possible gear, do not overuse the choke, disconnect your air conditioner, and so on. Now, it's true that if you follow such tips you and your car will be performing the function of conserving fuel, but it is worse than misleading - it is simply false - to claim that there is in the car a fuel-saving device. That is, there is no physical token—however complex—which corresponds to the functional description 'fuel-saver.' (Wilkes 1989a: 163, quoting Sheman)

Wilkes would want her money back whereas enactivists would not, although they might also lodge a complaint about the misleading term 'device.' The enactivist solution is not to look for or expect to find a device or physically structured object. The enactivist would be satisfied with what Wilkes had called tacit representations (at the personal level)—"dispositions, abilities, know-how: where and what we can do depends upon the way we are, or—sometimes—on what we have learned," to which we can add habits of avoiding jack rabbit starts, using the highest possible gear, not overusing the choke, disconnecting your air conditioner, and so on. Now Wilkes might say of the enactivist, 'a sucker is born every minute;' and the enactivist might reply, the sucker is the one who expected to find a fuel-saving device in the box. Although that's not Wilkes, she still seems to worry that there is no such device.

Second, Wilkes considers a thesis by Evan Macphail (1982, 1986). She summarizes:

> [His] hypothesis proposes that there is no quantitative or qualitative difference in intelligence among the vertebrate species, excluding man. He claims that there is no solid evidence that any of the cognitive feats ascribed to allegedly more intelligent species, like chimpanzees, cannot be rivalled by any other vertebrate—once one has taken into account and allowed for differ-

> ences in perceptual capacity, motivation, physical capacity, and other such contextual variables. (Wilkes 1989a: 178)

Wilkes suggests this is a radical claim that is difficult to access—she neither accepts nor rejects it, but finds it useful to make a point, which is about the importance of language.

Enactivists would reject Macphail's claim outright, not because they would reject the importance of language, but because it presumes to define intelligence as if intelligence were constitutionally independent of all the differences listed, which are primarily differences in embodiment, which also means differences in brain structure, motility, skills, etc.—all of which adds up to what we consider to be intelligence. Macphail's hypothesis would deny that intentions are embodied, embedded (in physical and social environments), enactive, affective, and perhaps extended. Rather, on Macphail's hypothesis, intentions are exclusively the product of human linguistic ability. Wilkes too points to the idea that the addition of language is what allows for the addition of intention formation:

> [I]f language is, as I believe, crucial for consideration of the need to postulate intentions, and if chimpanzees have some capacity for linguistic communication, then maybe some chimpanzee behaviour might require us to talk in terms of goal-representations. If not, not. (Wilkes 1989a: 180)

This would be to ignore P-intentions and M-intentions, but also the embodied and situated (wider) features of deliberation and D(istal)-intention formation, which are, at least in the human, always (explicitly or implicitly) socially embedded. Wilkes instead opts for a narrow conception of, if not in-the-head, then in-the-sentence form of intentionality:

> This is a very modest conclusion, though. It restricts 'goal representations' to language-using creatures, and even there argues for their utility only in cases where the deliberation, and framing of intentions, is explicit, prior to the action taken, and linguistic. (Wilkes 1989a: 182)

Consider, however, the rat. Wilkes considers a suggestion made by David McFarland, that "rats are capable of some cognitive evaluation" (McFarland 1989: 223; Wilkes 1989b: 209). McFarland, appealing to experiments by Adams and Dickinson (1981), offers what I think is a curious claim, that rats can cognitively evaluate in a way that involves a practical inference operating on a proposition-like form, which means that "the animal makes use of declarative representations in evaluating the likely consequences of its behaviour" (McFarland 1989: 223), but that this does not involve goal-directedness.

Wilkes clearly rejects McFarland's rejection of the goal-directed nature of such representations. I think that for Wilkes, *if* rats cognitively evaluate, then that involves intentionality and goal-directedness. Still, she does not necessarily accept that rats can cognitively evaluate (or deliberate). "I am left agnostic about 'representations of goals' in non-human animals, or human behaviours that do not obviously require linguistic talents" (Wilkes 1989b: 209).

More recent empirical evidence suggests that rats do deliberate about goals, although whether this involves propositional declarative representations is, to say the least, an open question. Martin Miłkowski summarizes some recent research:

> [R]odents plan future paths, which is reflected in the future-oriented navigational activity of place cells in the hippocampus in the brains of rats. This activity was directly observed in an elegant experiment (Pfeiffer and Foster [2013]). As it turns out, rats pause before taking a journey. During that pause, place cells emit sharp-wave-ripple events: irregular bursts of brief (100– 200 ms) large-amplitude and high-frequency (140–200 Hz) activity. These are distinct from regular spikes in place cells. Using an algorithm proposed earlier for decoding similar neural events […] Pfeiffer and Foster were able to show that place cells are used to represent the future journey of the rat to the location of a previously observed reward. (Miłkowski 2023: §5.2).

Milkowski is making a claim about subpersonal neural (structural rather than declarative) representations. Dennett might accept this, but enactivists and Wilkes would likely reject it. In any case this puts us right back into the problematic that Wilkes was wrestling with, and it suggests that we still have to work out some unresolved issues. In effect, Wilkes' analysis continues to be directly relevant to contemporary discussions.

References

Adams, C.D. and Dickinson, A. 1981. "Instrumental responding following reinforcer devaluation." *Quarterly Journal of Experimental Psychology* 33 (2b): 109–112.

Anscombe, G. E. M. 1957. *Intention*. Oxford: Blackwell.

Chisholm, R. M. 1957. *Perceiving: A Philosophical Study*. Ithaca and London: Cornell University Press.

Coles, M. G. 1989. "Modern mind-brain reading: psychophysiology, physiology, and cognition." *Psychophysiology* 26 (3): 251–269.

Cox, D. D., and Savoy, R. L. 2003. "Functional magnetic resonance imaging (fMRI)'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex." *Neuroimage* 19 (2): 261–270.

Dennett, D. C. 1989. "Comments." In Montefiore, A. and Noble, D. (eds.). *Goals, No-goals and Own Goals. A Debate on Goal-directed and Intentional Behaviour*. London: Unwin Hyman, 229–237.

Dennett, D. C. 1982-3. "Styles of mental representation." *Proceedings of Aristotelian Society* LXXXIII: 213–226.

Egan, F. 2014. "Explaining representation: a reply to Matthen." *Philosophical Studies* 170: 137–142.

Friston, K. J., and Frith, C. D. 2015. "Active inference, communication and hermeneutics." *Cortex* 68: 129–143.

Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale: Lawrence Erlbaum Associates.

Frith, C. D. and Gallagher, S. 2002. "Models of the pathological mind: An interview with Christopher Frith." *Journal of Consciousness Studies* 9 (4): 57–80.

Gabor, D. 1946. "Theory of communication. Part 1: The analysis of information." *Journal of the Institution of Electrical Engineers-part III: Radio and Communication Engineering* 93 (26): 429–441.

Gallagher, S. and Allen, M. 2018. "Active inference, enactivism and the hermeneutics of social cognition." *Synthese* 195 (6): 2627–2648.

Gerrans P. 2014. *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. Cambridge: MIT Press.

Haynes J-D, et al. 2007. "Reading hidden intentions in the human brain." *Current Biology* 17 (4): 323–328.

MacPhail, E. M. 1982. *Brain and Intelligence in Vertebrates*. Oxford: Clarendon Press.

MacPhail, E. M. 1986. "Vertebrate intelligence: the null hypothesis." In L. Weiskrantz (ed.). *Animal Intelligence*. Oxford: The Clarendon Press, 37–50.

McFarland, D. J. 1989. *Problems of Animal Behaviour*. London: Longmans.

Miłkowski, M. 2023. "Correspondence theory of semantic information." *The British Journal for the Philosophy of Science* 74 (2): 485–510.

Pacherie, E. 2008. "The phenomenology of action: A conceptual framework." *Cognition* 107 (1): 179–217.

Pfeiffer, B. E. and Foster, D. J. 2013. "Hippocampal place-cell sequences depict future paths to remembered goals." *Nature* 497: 74–79.

Piccinini, G. 2022. "Situated neural representations: Solving the problems of content." *Frontiers in Neurorobotics* 16: 846979.

Putnam, H. 1980. "Philosophy and our mental life." In N. Block (ed.). *Readings in Philosophy of Psychology*, Volume I. Cambridge: Harvard University Press, 134–143.

Schaffner, K. F. 2020. "Approaches to multilevel models of fear: The what, where, why, how, and how much?". In K. S. Kendler, J. Parnas, and P. Zachar (eds.). *Levels of analysis in psychopathology: Cross-disciplinary perspectives*. Cambridge University Press, 384–409.

Wilkes, K. 1989a. "Representations and explanations." In Montefiore, A. and Noble, D. (eds.). *Goals, No-goals and Own Goals. A Debate on Goal-directed and Intentional Behaviour*. London: Unwin Hyman, 159–184.

Wilkes, K. 1989b. "Explanations—How not to miss the point." In Montefiore, A. and Noble, D. (eds.). *Goals, No-goals and Own Goals. A Debate on Goal-directed and Intentional Behaviour*. London: Unwin Hyman, 194–210.