Final draft – please cite published version in J Robert Thompson (ed.) *The Routledge Handbook of Philosophy and Implicit Cognition.*
Routledge: forthcoming.

The fragmented mind: personal and subpersonal approaches to implicit mental states

Zoe Drayson

Abstract

In some situations, we attribute intentional mental states to a person despite their inability to articulate the contents in question: these are *implicit* mental states. Attributions of implicit mental states raise certain philosophical challenges related to rationality, concept possession, and privileged access. In the philosophical literature, there are two distinct strategies for addressing these challenges, depending on whether the content attributions are personal-level or subpersonal-level. This paper explores the difference between personal-level and subpersonal-level approaches to implicit mental state attribution and investigates the relationship between the two approaches. It concludes by highlighting the methodological and metaphilosophical commitments which can result in different perspectives on the relative priority of personal-level and subpersonal-level theories.

1. Implicit mental states

1.1 Attributions of implicit mental states

If I ask you what is on your mind, you might answer by telling me that you have noticed a cat on your porch, that you intend to find its owner, and that you want to keep the cat if nobody claims it. Philosophers tend to classify such mental states as *explicit*, meaning that the person can articulate the intentional contents of their thoughts by means of a sentence, given suitable prompting (see, for example Dummett 1991 and Davies 2015). Many of our everyday mental state attributions are of explicit mental states: the person can articulate the ascribed contents when prompted. In some situations, however, we attribute a mental state to a person who cannot suitably articulate the content in question. These are attributions of *implicit* mental states. Let us consider some examples of situations which might motivate us to attribute implicit mental states to people:

A. Skilled behavior. When a person exhibits certain skilled behavior, such as riding a bike or playing the piano, we often say that the person *knows how* to ride a bike or *knows how* play the piano. But the person is often unable to articulate the content of this practical knowledge: they cannot describe what it is they know. If these ascriptions of practical knowledge are genuine attributions of intentional mental states, then they must therefore be attributions of *implicit* mental states.

B. Closure of belief under logical consequence. If beliefs are closed under logical consequence, as models of epistemic logic suggest, then a person believes all the logical consequences of their explicit beliefs. Some of these logical consequences will be the contents of further explicit beliefs: someone who articulate the conjunctive belief *that Paris is the capital of France and home to the Louvre gallery* will generally also articulate a belief in the conjunct *that Paris is home to the Louvre gallery*. But some of the logical consequences of our explicit beliefs are not articulable due to real-world constraints on our time, deductive power, and working memory. If we are to be attributed beliefs in these non-articulable contents, such attributions must be of *implicit* mental states.[1]

C. Behavior/testimony mismatch. The beliefs we attribute to someone to make sense of what they say are often the same beliefs that would make sense of their behavior: the testimonial and predictive-explanatory roles of belief-attribution usually coincide (Schwitzgebel 2021). But there are cases where a person who sincerely voices an opinion also behaves in a way which suggests they endorse a contradictory opinion. Someone might act in a way that would be best explained by attributing a racist belief to

[1] Giordani (2015), for example, acknowledges that the simplest solution to problems of "logical omniscience" involves introducing a distinction between explicit and implicit belief and claiming that only the set of implicit beliefs is closed under logical consequence.

them, for example, but articulate and defend a contradictory anti-racist belief. If we are to explain the person's behavior in terms of their mental states, we might be motivated to attribute to the racist belief to them as an *implicit* mental state. (See Gendler 2018 for more examples and discussions of belief-discordant behavior.) In addition to these everyday cases, some pathological cases seem to motivate explanation in terms of implicit mental states: the performance of subjects with blindsight or visual form agnosia seem to be best explained by attributing information to them which is at odds with the beliefs that they articulate.

D. Visual perception. Both philosophical and scientific studies of perception have long suggested that people possess more visual information about the world than they can articulate. Evans (1982), for example, proposes that we can perceptually experience more shades of color than we can verbally describe or classify. And in vision science, it is standard practise to attribute the perceiver with information they cannot articulate (e.g. about complex probabilities or retinal disparity) to account for their perceptual capacities (e.g. to discriminate objects from their backgrounds or to make judgements of depth). Both cases involve attributing an *implicit* mental state to a person: an intentional content which the person cannot articulate.

1.2 Challenges for implicit mental state attribution

While it is commonplace to make attributions of implicit mental states in the above contexts, this practise faces several philosophical challenges. The very concept of an *implicit mental state* seems to challenge certain long-held assumptions about the nature of the mind. I'll consider three such assumptions here: rational conditions on mentality, constraints on concept possession, and special access which thinkers seem have to their own thoughts.

Rationality: Philosophers often emphasize the connection between mentality and rationality. Davidson (1980) famously proposes that attributions of propositional mental states only make sense against a background assumption of rational relations between thought and action: someone who believes that *p* rationally ought to behave as if *p* were true, and thus assert that p under the appropriate conditions.[2] If such rational conditions are a prerequisite for attributing intentional mental states, then attributions of *implicit* mental states are never appropriate.

Concepts: Assuming we *can* rationally attribute implicit mental states to people, there is still the question of which implicit mental states to attribute. It is widely held that there is a *conceptual constraint* on mental state attributions: when we specify the content of a thought, we should only employ concepts possessed by the thinker (Bermudez and Cahen 2020). And many philosophers (e.g. Evans 1982, Peacocke 1992, Heck 2007) propose that concept possession itself faces a further *generality constraint*: thoughts systematically connect to each other in virtue of their constituent concepts, and so the possessor of a concept must be able to utilize the concept in a variety of different thoughts.[3] Attributions of *implicit* mental states sometimes seem to violate these constraints. A vision scientist might explain a child's ability to perceive depth in terms of the information the possess about retinal disparity, even where the child is unable to employ the concept of retinal disparity in their thought more generally.

Privileged access: We often seem to have a certain special kind of epistemic access to our own mental states. A strong version of the privileged access claim, on which thinkers are omniscient or infallible with respect to the contents of their mental states,

[2] Davidson proposes that "if we are intelligibly to attribute attitudes and beliefs, or usefully to describe motions as behavior, then we are committed to finding, in the pattern of behavior, belief and desire, a large degree of rationality and consistency" (Davidson 1980, 237). See also Yalowitz's claim that "something only counts as being a mind—and thus an appropriate object of psychological attributions—if it meets up to certain rational standards" (Yalowitz 2005).

[3] Conceptual constraints are sometimes formulated as linguistic constraints: Davidson (1980) holds that propositional thought contents must be linguistically expressible by the thinker. See also Frege's claim that we can only grasp the content of a thought when it "clothes itself in the material garment of a sentence" (Frege 1956, 292).

> would presumably rule out implicit mental states completely.[4] Weaker forms of privileged access, according to which we merely have some sort of fallible access to certain of our mental states, might be consistent with attributions of implicit mental states, but they raise further questions about why some mental states are introspectable while others are not.

One way to respond to these three challenges would be to refrain from attributing implicit mental states. This approach would result in some cases of seemingly rational action being redescribed as reflex-like behavior, and a loss of the ability to distinguish intellectual capacities from bodily abilities. I will set aside such an approach here to focus instead on how philosophers have attempted to retain attributions of implicit mental states in a way which can be reconciled with the challenges outlined above. I will propose that there are two distinct ways to attribute implicit mental states, each of which employs different strategies for addressing the challenges of rationality, concept possession, and privileged access.

1.3 Personal and subpersonal attributions of implicit mental states

Traditionally, mental state attributions ascribe intentional content to the person or thinker as a whole. When we say that a person represents that *p*, calculates that *q*, or predicts that *r*, for example, we are describing the *person* as grasping a propositional thought. These are *personal-level* attributions of mental states. Since the birth of computational cognitive psychology, however, it has become common to make attributions of intentional content below the level of the person. When we describe a neural structure as representing that *p*, the visual system as calculating that *q*, or a Bayesian network as predicting that *r*, for example, we are attributing an intentional content to some proper part of the thinker, such a functional subsystem or a representational vehicle. These are *subpersonal-level* attributions of mental states. The distinction here between *personal-level* and *subpersonal-level* approaches is first and foremost

[4] Examples of such strong versions of privileged access include claims of self-intimation, self-presentation, or luminosity: being in a mental state suffices for knowing that one is in that mental state. Foundationalist epistemologies often rely on such privileged access claims.

a distinction between two ways of theorizing about the mind, which need not be understood as competing theories. I will consider questions about the semantic, epistemic, and ontological interpretation of these theories, and the relationship between personal-level and subpersonal level theories, in Section 4. (In this paper I will be focusing on personal-level and subpersonal-level approaches to attributing *implicit* mental states, but both approaches can arguably also be used to attribute *explicit* mental states.[5])

To get an idea of how personal-level and subpersonal-level approaches can be applied to attributions of implicit mental states, consider a case of skilled behavior such as my ability to play 'Moon River' on the piano. Both approaches make attributions of intentional content, but in a different way. For an example of a *personal-level* approach, consider how Stanley and Williamson (2001) account for skilled behavior in terms of the *person* standing in a knowledge relation to a propositional content. In this case, I know *that w is the way to play 'Moon River' on the piano*, where *w* is a proposition which I cannot articulate. For an example of a *subpersonal-level* approach, on the other hand, consider how Fodor (1968) accounts for skilled behavior as the competence of an information-processing system. He compares the person's ability (e.g. to play the piano) with a computer's ability to calculate: intentional contents in the form of "propositions, maxims, or instructions" (Fodor 1968, 638) are attributed to proper parts of the of the information-processing system (vehicles of representation) rather than to the person as a whole.

In what follows, I'll explore the personal-level and subpersonal-level approaches to implicit mental state attribution in more detail and consider further examples. I will show that personal-level and subpersonal-level approaches have very different strategies for addressing the

[5] Fodor's (1975) 'Language of Thought' hypothesis, for example, is a subpersonal-level theory of information-processing systems on which computational states can correspond to either explicit or implicit mental states. See also Stich's (1978) claim that *explicit* beliefs can be understood in terms of cognitive *subsystems*.

challenges faced by implicit mental state attribution concerning rationality, concept possession, and privileged access.

2. Personal-level approaches to implicit mental states

Personal-level attributions of implicit mental states describe a person as grasping an intentional mental content, and yet not being in a position to assert it. How does the personal-level theorist address the apparent irrationality of such an attribution? Consider, first, the piano-playing example outlined above. A personal-level theorist might appeal to different ways of grasping a content, some of which are rationally compatible with being unable to articulate the content. For Stanley and Williamson (2001), for example, knowing how to φ involves a proposition being presented to the person in a practical rather than theoretical way: it is a case of knowing that *w* is a contextually relevant way to φ under a practical mode of presentation. This allows the personal-level theorist to reconcile constraints on mental state attribution with the person's inability to linguistically articulate the proposition involved.

This approach, however, doesn't seem to easily extend to other cases where we want to attribute implicit mental states. What if the implicit beliefs we attribute to the person directly contradict their explicit beliefs? Consider Lewis's (1982) example of someone who explicitly believes that Nassau Street runs roughly east-west; explicitly believes that the railroad nearby runs roughly north-south; and explicitly believes that the two are roughly parallel. Any two of these propositions entail the negation of the third, so closure under logical consequence requires that we attribute a contradictory (presumably implicit) belief to the person. The fact that the contradictory belief is attributed as an *implicit* mental state does not make the person any more rational: attributions of contradictory belief violate even the most basic constraints on rationality, leaving no justification for ascribing any intentional mental states at all.[6] One

[6] Davidson argues that "[n]othing a person could say or do would count as good enough grounds for the attribution of a straightforwardly and obviously contradictory belief" (Davidson 1985, 138). If propositions are sets of possible worlds, contradictory beliefs would require impossible worlds.

way a personal-level theorist might address this problem is to relativize belief attributions to temporal stages (time slices) of the person. Lewis himself takes this approach, suggesting that "the blatantly inconsistent conjunction of the three sentences […] was not true according to be beliefs":

> "My system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once." (Lewis 1982, 436)

Lewis's solution is to propose that we understand minds as temporally *fragmented*: a person's grasp of a thought is always relativized to a particular time. If the three beliefs are not *simultaneously* attributable to the person, then the conditions for logical consequence are never met and we do not need to attribute the implicit (contradictory) belief in the first place. Lewis's strategy of temporally-relativized belief attributions might also be applied to behavior/testimony mismatch cases: if we deny any overlap between the person-stage who (explicitly) believes *that p* and the person stage who (implicitly) believes *that not-p*, there is no person-stage to whom we must attribute the belief *that p and not-p*.

In some examples of implicit mental state attribution, where we are motivated to attribute the belief that *p* and the belief that *not-p* to a person simultaneously, the temporal fragmentation story will not help the personal-level theorist. Consider a visual perception case, for example. To explain why someone is subject to an optical illusion, we might say that the person implicitly believes that light is coming from above while they *simultaneously* articulate the belief that light is coming from below. In order to preserve the rationality constraint on mental state attributions, the personal-level theorist might relativize such attributions to tasks, contexts, or purposes, as well as times. Varieties of this approach in the literature include Egan's (2008) suggestion that different beliefs drive different aspects of our behavior in different contexts; and Elga and Rayo's (2021) proposal that explaining and predicting behavior requires specifying what information is available to an agent relative to various purposes. In the visual perception case, the personal-level theorist could argue that the person *believes-for-a-vision-guiding-*

purpose that light is coming from above and *believes-for-a-more-general-purpose* that light is coming from below, where only the latter purpose allows articulation of the belief content.[7]

Even if we allow that these personal-level approaches to implicit mental state attribution preserve the rationality constraint on mental state attribution, there is still the question of which content to attribute. Some approaches to content interpretation (e.g. Davidson 1982) seem to require the sort of rational holism which is denied by relativizing mental state attributions to temporal or contextual fragments.[8] If a personal-level theorist argues that we can relativize concept possession and privileged access to times, contexts, or purposes, this seems to violate the generality constraint: Evans (1982) concludes that someone who can't see the connection between thoughts involving the same concepts cannot really be said to grasp either of the thoughts in question. Stanley, however, proposes that at least in the skilled behavior case, the inability to linguistically articulate the content of one's mental state is compatible with being able to conceptualize, introspect, and perhaps even assert the content in a demonstrative form: e.g. "I know that *this* is the way to play Moon River on the piano".

3. Subpersonal-level approaches to implicit mental states

Subpersonal-level theories start from the assumption that minds can be understood as information-processing systems, and that the capacities of an information-processing system can be functionally decomposed into informational subsystems and discrete computational states:

> "Sub-personal theories proceed by analyzing a person into an organization of subsystems […] and attempting to explain the behaviour of the whole person as the outcome of the interaction of these subsystems" (Dennett 1978, 154).

[7] For another approach to personal-level theorizing about implicit mental states, not discussed here, see Schwitzgebel's (2001) account of "in-between believing".

[8] See also Gozzano (1999) for the idea that such mental fragmentation strategies avoid irrationality only by introducing more complex problems.

What makes these subpersonal-level psychological theories rather than non-psychological descriptions of physical mechanisms is the intentional interpretation of the subsystems. Subpersonal-level theories describe proper parts of the system (not just whole persons) as representing, evaluating, calculating, expecting, discriminating, and so on. (I'll address the standard worries about this practise shortly.) Such theories end up attributing contents to subpersonal vehicles of representation: symbols, clusters in state spaces, or attractor basins, depending on the sorts of computational architectures we are dealing with. Importantly, these subpersonal vehicles of representation also have *non-semantic* properties in virtue of which they can physically implement computational transitions and be assigned contents in a naturalistic way. [9] This means that cognitive psychology has at least one way to theorize about representational vehicles which does not characterize them in normatively-constrained terms. Considered non-semantically, there is no expectation that these vehicles meet semantic constraints on rationality or concept-possession.[10] There is no assumption that subpersonal vehicles are introspectable, and whether we have privileged access to their contents will depend on the nature of the information-processing architecture.

The challenges faced by implicit mental state attribution, discussion in Section 1.2, are largely semantic or normative. Where personal-level approaches attempt to show how implicit mental state attributions are compatible with constraints on rationality and concept possession, subpersonal-level approaches suggest that these challenges do not arise in the first place. To see how this works, let us return to the piano-playing example. Cognitive psychology characterizes people's skilled abilities partly in terms of information-processing performed by their motor-control subsystem. Subpersonal-level theories attribute contents (concerning the aims of our movements and how to achieve them, for example) to motor representations, where these can be characterized in terms of their non-semantic vehicle properties and thus

[9] For more on the nature representational vehicles and the importance of the distinction between their semantic and non-semantic properties, see Drayson (2018).

[10] Of course, there may still be *syntactic* constraints on cognitive subsystems: we might understand the physical system (e.g. the brain) as implementing a formal language.

individuated without appeal to rational norms. If we adopt a naturalistic theory of content determination, we can describe the motor representations as carrying information about the biomechanical constraints and kinematic rules relevant to piano-playing: information about the mathematical relationship between the velocity and amplitude of movement, for example, or laws relating curvature and velocity (see Mylopoulos and Pacherie (2017) for further discussion). Such attributions of content are compatible with the person being unable to reason more generally about these physical and mathematical constraints, and lacking first-person privileged access to these contents. We might account for the lack of articulability of the content in question in terms of the computational architecture of the motor subsystem: perhaps it works independently from other cognitive subsystems which we can introspect, or perhaps motor representations are carried in a different format to those representations which we can articulate.

A similar approach can be applied to some of the other examples of implicit mental state attributions outlined in Section 1.1. In the perception case, subpersonal-level theorists can attribute assumptions about retinal disparity to the visual system rather than to the person. One might explain the non-introspectability of early visual processing by positing that low-level perceptual processes use a different representational format from other cognition (e.g. iconic versus discursive), or that they use a different kind of computational processing from other cognition (e.g. connectionist versus classical).[11] Notice that subpersonal-level theories tend to be engaged in an explanatory psychological project rather than a justificatory epistemic project: subpersonal theories of visual perception, for example, are not trying to address skeptical worries, but to explain how our perceptual mechanisms operate. Similarly, since subpersonal-level theorists are generally interested in describing our inferential thought processes rather

[11] For examples of some of the different computational ways to account for non-introspectable psychological states, see Fodor (1983) on informationally encapsulated modules, Frankish (2010) on type-1 and type-2 processes, and Hohwy (2013) on statistical boundaries in hierarchical Bayesian architectures.

than justifying them, the normative epistemic problems raised by closure under logical consequence are ones they can set aside.

Subpersonal-level approaches have faced challenges of their own, however. When cognitive psychologists first started talking about computational states as 'representing' states of affairs, some philosophers were skeptical: how can anything other than a person genuinely represent the world as being a certain way?[12] There were two prominent criticisms of subpersonal-level approaches. First, does it even make sense to apply mental terminology below the level of the person, or does applying predicates true of the whole person to one of its proper parts commit a "mereological fallacy"? And second, even if mental terminology could be applied below the level of the person, wouldn't any attempt to explain contentful systems in terms of contentful parts lead to some sort of "homuncular regress"? But subpersonal-level approaches have been responsible for much of the success of cognitive science, acting as a bridge between traditional personal-level approaches to the mind and lower-level neurophysiological explanation. This has prompted many philosophers to rethink their criticisms and to reject or at least reconsider these challenges: it is not obviously wrong to apply a psychological predicate to a cognitive subsystem, and such attributions may eventually 'bottom-out' rather than generate regress worries.[13]

4. The relationship between personal-level and subpersonal-level approaches to implicit mental states

We have seen that there are two ways to attribute intentional mental content: personal-level approaches and subpersonal-level approaches. It is important to remember, however, that

[12] See Stich's acknowledgement, for example, that many philosophers including himself were "skeptical about the idea of invoking internal representations in psychological theories" in the early days of subpersonal-level psychology (Stich 2011, xix).

[13] For more on the mereological and homunculus fallacies and their proposed resolution, see Drayson (2012, 2014, 2017).

these are not mutually exclusive ways of theorizing about the mind. Many philosophers of cognitive science propose that personal-level theories can be complemented, expanded upon, and even vindicated by subpersonal-level theories.[14] To see how this might work with attributions of *implicit* mental states, consider the following examples which aim to combine personal-level and subpersonal-theories theories.

> Skilled behavior: A personal-level theory of skilled behavior can be combined with a subpersonal-level theory of motor control from cognitive science. Pavese (2017) proposes that by attributing intentional content to representational vehicles in the motor subsystem, we can give a more rigorous characterization of Stanley and Williamson's (2001) notion of a practical mode of presentation.
>
> Testimony/behavior mismatch: The fragmentation approach to personal-level theorizing, which attributes beliefs to a person relative to context or task, can be further cashed out in terms of a subpersonal-level theory of cognitive architecture. Bendaña and Mandelbaum (2021), for example, account for personal-level fragmentation by attributing informational contents to distinct and functionally isolated data structures within the cognitive architecture, instead of one single database for information storage.

If these strategies work, then we do not have to choose between personal-level and subpersonal-level theorizing: there are situations in which we can attribute both personal-level and subpersonal-level intentional contents.

[14] Notable examples includes Fodor's suggestion that "having a particular propositional attitude is being in some computational relation to an internal representation" (Fodor 1975, 198) and Lycan's proposal that the posits of personal-level theories can be identified with "the property of having such-and-such an institutionally characterized state of affairs obtaining in one (or more) of one's appropriate homunctional departments or subagencies" (Lycan 1988, 41). Davies summarizes such approaches as follows: "we assume that, if a person consciously or occurrently thinks that p, then there is a state that has the representational content that p and is of a type that can figure in subpersonal-level psychological structures and processes" (Davies 2005, 370).

Whether strategies like these work is a matter for further investigation. If our best subpersonal-level theories of motor representation attribute contents which the person does not conceptually possess and reject a role for contents in reference determination, then motor representations do not seem to be able to play the role attributed by Stanley and Williamson's (2001) to practical modes of presentation.[15] And what if the personal-level attributions of fragmented beliefs which account for the mismatch between testimony and behavior cross-cut the sorts of informational divisions proposed by our empirically well-founded subpersonal theories of cognitive architecture?[16] In these less straightforward cases, combining personal-level and subpersonal-level theories will require making adjustments to one theory or the other. Philosophical opinions, I will suggest, differ widely on whether we should focus on adjusting our person-level theories or our subpersonal-level theories.

Some philosophers propose that we ought to prioritize the personal-level approach to intentional content attributions. If the subpersonal level theory doesn't support the personal level theory, they suggest that we should replace it with a different subpersonal theory, or perhaps even deny the need for subpersonal theorizing. According to this view, we can accept that a person represents that *p* in a fragmented way, without thinking this requires that some architectural 'fragment' of the person represents that *p*. Schwitzgebel appears to be endorsing this approach to implicit mental states when he suggests that "[w]e can accept disunity [personal-level fragmentation] without embracing the dubious architectural commitments of system [subpersonal-level architectural] fragmentation" (Schwitzgebel 2021, 368). Some philosophers take the opposite approach and argue that we should prioritize the subpersonal-level theory of intentional content attributions. If our best subpersonal-level theory is at odds with our personal-level theory, they propose that our personal-level theory is the one which

[15] For a development of this argument along these lines, see Schwartz and Drayson (2019).

[16] Norby (2014) proposes that the sorts of fragmented belief attributions we make in some personal level theories do indeed cut across the division drawn by empirical psychology. Schwitzgebel (2021) concurs that while certain kinds of personal-level fragmentation stories may find empirical support, there other cases of implicit mental state attribution which don't match up neatly with the empirically well-founded varieties of subpersonal-level explanation.

ought to be reformulated or even rejected. Proponents of this approach to implicit mental states (e.g. Norby 2014) argue that if empirical psychology divides up psychological state types in a way unsuited to a personal-level fragmentation theories, then this is a problem for personal-level theory. In what follows, I will suggest that the two sides in this debate are motivated by very different methodological and metaphilosophical views.

Philosophers who prioritize personal-level approaches often propose that such approaches rely on distinctive epistemic methods (e.g. rational reflection, introspection, intuition, conceptual analysis, transcendental reasoning) which they take to be more reliable than the scientific reasoning that results in our subpersonal-level theories. Some (e.g. McDowell 1994) go further, proposing that subpersonal theories are irrelevant to our metaphysical understanding of the mind because they take personal-level theories to be governed by normative principles of rationality which make them completely autonomous from subpersonal-level theories.[17] On such views, realism about personal-level theories does not need to be supported or vindicated by subpersonal-level theories: semantic facts don't need to explained by non-semantic facts; *abstracta* don't need to be explained in terms of *concreta*. Proponents of prioritizing the personal-level approach often seem to think that only personal-level theorizing can provide genuine metaphysical insight into the fundamental nature of the mind.[18]

Philosophers who prioritize subpersonal-level approaches, on the other hand, are often motivated by naturalistic concerns about some of the methods associated with personal-level theorizing. They tend to worry about the status of analyticity, the reliability of intuition, and the

[17] This is what Rupert (2018) terms the "Received View" on which personal level facts are known non-scientifically by a priori reasoning, conceptual analysis, and introspection, while cognitive science has a more modest role studying the mere implementation of these facts.

[18] In the perception literature, for example, Logue follows McDowell in stating that the fundamental metaphysical structure is "that which provides the ultimate personal-level psychological explanation of the phenomenal, epistemological and behavioural facts." (Logue 2012, 212). Logue contrasts personal-level theories such with scientific theories which appeal to "subpersonal psychological facts (e.g. the perceptual processing in the brain that takes place between stimulation of the sensory organs and experience)" (Logue 2012, 212). For a counter-argument, see Drayson (2021).

possibility of *a priori* knowledge, and instead favor applying scientific reasoning methods to philosophy. Such philosophers (e.g. Churchland 1986) propose that if our best science attributes subpersonal-level content to internal vehicles of information processing, we should take this more seriously than our intuitive folk-psychological frameworks which attribute content to the person. Prioritizing the subpersonal-level theory in this way might lead us to reject realist interpretations of personal-level theorizing altogether, and even to think that subpersonal-level theorizing can give us *a posteriori* access to metaphysical truths about the mind.[19]

Not all philosophers of mind fall into one or other of these camps: we do not have to insist on a realist interpretation of either personal-level or subpersonal-level theories. Perhaps all our attributions of content, whether personal-level or subpersonal-level, are nothing more than heuristic tools. Or perhaps some of our theories are true in a deflationary sense which is not ontologically committing: we can accept them without believing in the entities they posit. (See Drayson 2022 for further discussion of the varieties of anti-realism in philosophy of mind.)

5. Conclusion

Some cases of intelligent behaviour motivate us to attribute intentional mental states to a person even where they are unable to articulate the intentional content in question. In such cases where we cannot readily attribute *explicit* mental states, we instead tend to make attributions of *implicit* mental states. But many of our traditional ideas about mental states – that mental state attributions can only be made against a background assumption of rationality, that the contents of mental states must be specifiable in concepts possessed by the person, and

[19] There are different arguments in this vicinity which can lead to an eliminativist conclusion. Churchland (1986) claims that none of our neural properties (action potentials, spreading activation, spiking frequencies, etc.) has the appropriate syntactically-structured features to vindicate folk-psychological theorizing; while Stich (1983) claims that even if we could make sense of syntactically-structured neural states, we couldn't individuate these structures in the same way that we individuate beliefs in our folk psychological discourse. Rupert (2018) suggests that the success of scientific explanations at the subpersonal level gives us reason *not* to posit anything essentially normative, rational, and person-level.

that the person has privileged access to their intentional mental contents – are difficult to reconcile with attributions of implicit mental states.

How we respond to this challenge will depend upon whether we take a personal-level approach or a subpersonal-level approach to attributions of implicit mental states. Personal-level approaches to implicit mental states ascribe intentional content to the person as a whole; while subpersonal-level approaches to implicit mental states ascribe intentional content to cognitive subsystems of the person. Each approach has different ways of addressing the challenges associated with rationality, concept possession and privileged access. Personal-level theories find ways to meet the conditions in question by relativizing mental state attributions to persons at a time or a context, for example, or invoking practical modes of presentation. Subpersonal-level theories tend to relax or reject the conditions in question, engaging in a description of psychological mechanisms with more minimal normative constraints.

It is tempting to think that the sort of rational fragmentation proposed by personal-level theorists must map on to the sort of informational fragmentation which we find in different models of subpersonal-level cognitive architecture. While this is one possibility, it is important to remember that the relationship between personal-level theorizing and subpersonal-level theorizing can be characterized in a variety of different ways, depending on one's background methodological and metaphilosophical views.

Acknowledgements

Earlier versions of this work benefited from discussions at the University of Nevada, Reno (2016) and the University of Arizona (2019). It was also presented at the following conferences: Implicit Attitudes, Explicit Attitudes and Join Action, Ruhr Universitat Bochum (2018); The Self and Its Realization, University of Colorado, Boulder (2018), and the annual meeting of the Southern Society for Philosophy and Psychology, San Antonio (2018). I am grateful to the graduate students of the Philosophy of Mind reading group at the University of California, Davis, for their feedback, and to the volume editor and an anonymous reviewer for helpful comments.

References

Bendaña, Joseph & Mandelbaum, Eric (2021). The Fragmentation of Belief. In Cristina Borgoni, Dirk Kindermann & Andrea Onofri (eds.), *The Fragmented Mind* (78-107). Oxford, UK

Bermúdez, José and Arnon Cahen, 'Nonconceptual Mental Content', *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.)

Churchland, P (1986). *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.

Davidson, Donald (1980). *Essays on Actions and Events*. Oxford: OUP.

Davidson, Donald (1982). Paradoxes of irrationality. In Richard Wollheim & James Hopkins (eds.), *Philosophical Essays on Freud* (289-305). Cambridge: Cambridge University Press.

Davidson, Donald (1985). Deception and division. In Lepore, Ernest & McLaughlin, Brian P. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (138-148). Oxford: Blackwell.

Davies, Martin (2005). Cognitive science. In Frank Jackson & Michael Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. New York: Oxford University Press New York.

Davies, M. (2015) Knowledge (explicit and implicit): philosophical aspects. *International Encyclopedia of the Social and Behavioral Sciences* (2nd edition), 74-90. Academic Press.

Dennett, Daniel C. (1978). *Brainstorms*. MIT Press.

Drayson, Zoe (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives* 26 (1):1-18.

Drayson, Zoe (2014). The Personal/Subpersonal Distinction. *Philosophy Compass* 9 (5):338-346.

Drayson, Zoe. (2017). Psychology, personal and subpersonal. In The Routledge Encyclopedia of Philosophy. Taylor and Francis. https://www.rep.routledge.com/articles/thematic/psychology-personal-and-subpersonal/v-1. doi:10.4324/9780415249126-V044-1

Drayson, Zoe (2018). The realizers and vehicles of mental representation. *Studies in History and Philosophy of Science Part A* 68:80-87.

Drayson, Zoe (2021). Naturalism and the metaphysics of perception. In Heather Logue & Louise Richardson (eds.), *Purpose and procedure in philosophy of perception* (215-233). Oxford University Press.

Drayson, Zoe (2022). What we talk about when we talk about mental states. In Tamas Demeter, Ted Parent and Adam Toon (eds.) *Mental Fictionalism: Philosophical Explorations* (147-159). Routledge.

Dummett, Michael (1991). *The Logical Basis of Metaphysics*. Harvard University Press (Cambridge, MA).

Egan, Andy (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies* 140 (1):47 - 63.

Elga, Adam & Rayo, Agustin (2021) Fragmentation and information access. In Cristina Borgoni, Dirk Kindermann and Andrea Onofri (eds), *The Fragmented Mind* (37-53). Oxford University Press.

Evans, Gareth (1982). *The Varieties of Reference*, edited by John McDowell. Oxford: Clarendon Press.

Fodor, Jerry A. (1968). The appeal to tacit knowledge in psychological explanation. Journal of Philosophy 65 (October):627-40.

Fodor, Jerry A. (1975). *Language of Thought.* MIT Press

Fodor, Jerry A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Frege, Gottlob (1956). The thought: A logical inquiry. *Mind* 65 (259):289-311.

Frankish, Keith (2010). Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass* 5 (10):914-926.

Gendler, Tamar Szabó (2008). Alief and Belief. *Journal of Philosophy* 105 (10):634-663.

Helmholtz, Hermann (1878). The Facts in Perception. In R. Kahl (ed.), *Selected Writings of Hermann Helmholtz*. Wesleyan University Press.

Giordani, Alessandro (2015). A suitable semantics for implicit and explicit belief. *Logique Et Analyse* 58 (231).

Gozzano, Simone (1999). Davidson on Rationality and Irrationality. In Mario De Caro (ed.), *Interpretations and Causes. New Perspectives on Donald Davidson's Philosophy*. Dordrecht: Synthese Library, Kluwer.

Hohwy, Jakob (2013). *The Predictive Mind*. Oxford University Press UK.

Lewis, David K. (1982). Logic for equivocators. *Noûs* 16 (3):431-441.

Logue, Heather (2012). Why Naive Realism? *Proceedings of the Aristotelian Society* 112 (2pt2):211-237.

Lycan, William G. (1988). *Judgement and Justification*. Cambridge University Press.

McDowell, John (1994). The content of perceptual experience. *Philosopical Quarterly* 44 (175):190-205.

Mylopoulos, Myrto, and Pacherie, Elisabeth (2017). Intentions and Motor Representations: the Interface Challenge, in *Review of Philosophy and Psychology*. 8(2): 317-336.

Norby, Aaron (2014). Against Fragmentation. *Thought: A Journal of Philosophy* 3 (1):30-38.

Pavese, Carlotta (2017). A Theory of Practical Meaning. *Philosophical Topics* 45 (2):65-96.

Peacocke, Christopher (1992). *A Study of Concepts*. MIT Press.

Rupert, Robert (2018) The Self in the Age of Cognitive Science: Decoupling the Self from the Personal Level. *Philosophic Exchange*, 47 (1): 1-36.

Schwartz, Arieh & Drayson, Zoe (2019). Intellectualism and the argument from cognitive science. *Philosophical Psychology* 32 (5):662-692.

Schwitzgebel, Eric (2001). In-between believing. *Philosophical Quarterly* 51 (202):76-82.

Schwitzgebel, Eric (2021). The pragmatic metaphysics of belief. In Cristina Borgoni, Dirk Kindermann and Andrea Onofri (eds), *The Fragmented Mind* (350-376). Oxford University Press.

Stanley, Jason (2011) *Know How* (Oxford U.K.: Oxford U.P.)

Stanley, Jason and Williamson, Timothy (2001) Knowing How. Journal of Philosophy, 98:8, pp. 411-444.

Stich, Stephen P. (1978). Beliefs and subdoxastic states. *Philosophy of Science* 45 (December):499-518.

Stich, Stephen P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. MIT Press.

Stich, Stephen (2011). *Collected Papers, Volume 1: Mind and Language, 1972-2010*. Oup Usa.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.

Yalowitz, Steven (2005). Anomalous monism. *Stanford Encyclopedia of Philosophy*.