Low-level Simulation, Unconscious Inference, and Implicit Belief

How can mental state types cross levels?

Section 1: Level-crossing mental-state types

I am interested in the idea that:
i) mental states can be both 'low-level' and 'high-level', and
ii) instances of both can belong to the same familiar mental-state type.

Examples of familiar mental-state types: inference, belief, knowledge, desire, imagination, perception, pride, doubt, etc.

By 'high-level/ low-level' I mean to gesture at a cluster of distinctions, not always well-defined (cf. Block 1995, Drayson 2012, Holroyd 2016), sometimes expressed using the terms:
- 'conscious/unconscious'
- 'personal/subpersonal'
- 'explicit/implicit'

Section 1: Level-crossing mental-state-types

My interest in this originates in a particular debate about social cognition.

Many simulation-theorists hold that much social cognition is a matter of 'implicit/low-level simulation', often associated with the mirror-neurone system (Gallese and Goldman 1998, Goldman and Sripada 2005, Goldman 2006, Michael et al. 2014). This would contrast with 'explicit/high-level simulation', which involves the conscious decision to 'project ourselves into' another's perspective.

One persistent line of criticism has denied, not the mirror neurone system underlies social cognition, but that its operations can be meaningfully called 'simulations' in the same sense as explicit/high-level simulations (Jacob 2002, Gallagher 2007, 2015, cf. Slors 2010).

That is, the coherence of simulation theory is in part a matter of whether there is a meaningful category of 'mental simulation' that crosses levels.

Section 1: Level-crossing mental-state-types

E.g. Gallagher writes:

*"...deciding that mirror neurons function as simulations depends on taking a...
model that was developed at the explicit, conscious, or personal level, and looking
for that... model at the neuronal level. But what precisely justifies this
interpretation?
...What theorists of implicit simulation (Gallese, Jeannerod, Pacherie), and even
critics of implicit ST, like Saxe (2005), call 'simulation'. is not simulation in any
genuine sense of the word."*

(Gallagher 2007, p.359)

Section 1: Level-crossing mental-state-types

Similar issues arise in many other areas:

- Does perceptual processing involve 'unconscious inferences'?
- Do humans have 'tacit knowledge' of grammatical rules, of 'folk psychology', of 'folk physics', etc.?
- What is the appropriate response to finding evidence that you are 'implicitly biased'? (Saul 2013, Holroyd 2014, Machery 2016, Mandelbaum 2016)

In each case, a category defined with reference to 'high-level' states ('inference', 'knowledge', 'bias', etc.) is extended to 'low-level' versions. In each case, there's room for doubt about whether those low-level versions really qualify (are unconscious inferences really inferences? Is implicit bias really bias?).

Section 2: The shape of the problem

I won't be able to say anything directly about any of these debates in particular.

But I think it is useful to look at the general shape of this kind of debate, and the sort of moves that are available.

The basic problem is what I'll call 'Incompatible Criteria': some of the criteria for being low-level conflict with some of the criteria for belonging to a given state type.

This means that for any specific sense of 'low-level' and specific mental-state type, it can't be taken for granted that low-level versions of that type are possible.

Section 2: The shape of the problem

So first, what does it mean to be 'low-level'? Different writers use different criteria, but standard themes include:

- **Control:** Low-level states may be in some sense 'involuntary', under reduced voluntary control, or just less dependent on conscious effort.
- **Integration:** Low-level states may be 'isolated' from other mental states, so that they are not affected by other states which ought to conflict with or remove them.
- **Access:** Low-level states may be 'inaccessible' to their subject, or harder to access, or require more effort to access.

Section 2: The shape of the problem

Next, what features could be used to categorise a mental-state token as belonging to a particular type?

- **Functional role:** Does it have the typical causes and effects, both in the environment and among other mental states, of that type?
- **Normative role:** Is it subject to the characteristic standards of success, failure, and malfunction of that type?
- **Phenomenology:** Does undergoing it feel the characteristic way that states of that type feel?
- **Content:** Does it represent the right kind of thing, in the right kind of way, to be a state of that type?

Section 2: The shape of the problem

Problems for categorisation by functional role:

How can high-level and low-level states have the same causes and effects, if one is voluntary, integrated, and accessible, and the other is involuntary, isolated, and inaccessible? Those are three significant differences in which other mental states they cause and are caused by.

Problems for categorisation by normative role:

How can low-level states be subject to the same normative standards as high-level ones, if their subjects can't control them and don't know about them? Generally, lack of control and knowledge undercuts most sorts of normative judgement.

Section 2: The shape of the problem

Problems for categorisation by phenomenology:

How can we *tell* that a low-level state has the same phenomenal character as a high-level state, if its subject can't access it? That seems to cut off the standard way of determining phenomenal character.

Problems for categorisation by content:

There's no problem with assigning low-level states a kind of 'content' based on which stimuli they seem to co-vary with, but that's not enough to assign them to a contentful mental-state type. Usually that assignment turns on features like functional role, normative role, and phenomenology - taking us back to the above problems.

Section 2: The shape of the problem

So what?

Because there are different ways to be 'low-level', and different ways to define mental-state types, these problems don't arise in any unitary way for every attempt to extend a familiar mental-state type across levels.

But they do support *prima facie* doubts about such attempts: we can't assume by default that any particular state type can have both 'high-level' and 'low-level' versions.

The burden is on whoever proposes 'low-level versions of' the familiar type, to show that they avoid the 'problem of Incompatible Criteria'.

Section 3: Some general solution-types

Distinguish two sorts of response to the problem of Incompatible Criteria:

Primary Responses: Argue that the criteria for being low-level in the specific sense you intend, and the criteria for belonging to the specific mental-state type you are discussing, don't conflict.

Secondary Responses: Accept that the criteria conflict, and that low-level states in the relevant sense probably *can't* belong to the familiar mental-state type. Then make some sort of adjustment to your claims that removes this problem.

Section 3: Some general solution-types

Primary responses are very specific to particular debates.

E.g. If 'low-level' is defined specifically in epistemic terms (can't be accessed introspectively), and 'belief' is defined specifically in non-epistemic functional terms (e.g. by relationships to evidence, other beliefs, emotion, and action), then it is consistent to talk about 'low-level ('unconscious') beliefs.'

Or if 'low-level' is defined by automaticity rather than access, low-level states can be assigned to phenomenologically-defined categories (e.g. emotions) but not to those defined by their voluntariness (intentions).

Section 3: Some general solution-types

Three common sorts of secondary response:

1. Use a new definition
2. Relativise the states to a sub-system
3. Appeal to level-crossing tokens

Section 3: Some general solution-types

1. Use a new definition:

Abandon the familiar mental-state concept in favour of a new, theoretically-motivated definition tailored to cover both high-level and low-level states.

This might mean inventing a new term, like 'alief' (Gendler 2008a, 2008b) or 'subdoxastic states' (Stich 1978), that contrasts with the familiar term (e.g. belief) but is adjacent.

Or it might mean offering formal definitions of a familiar term which are not tethered to everyday usage (this is one way to read of the simulation debate, e.g. Hurley 2008, Goldman 2008).

Section 3: Some general solution-types

2. Relativise the states to a sub-system

Perhaps a low-level state has a functional role that doesn't qualify as that of a particular mental state, relative to all the other states of that subject, but *do* qualify relative to some important subset of them.

E.g. 'I don't know the rules of English grammar, but my language module does!' = there are states in me that play the belief-role relative to the inputs, outputs, and other states assigned to 'the language model'.

I think this manner of speaking may be widespread in everyday idioms ('I know it's safe, but my gut doesn't', or 'I know it's this way, but my brain keeps trying to see it that way').

Section 3: Some general solution-types

3. Appeal to level-crossing tokens:

Persisting mental states seem to sometimes go from 'high-level' to 'low-level' by automatisation (over time they become more automatic and less accessible), or from 'low-level' to 'high-level' by articulation (with attention or training they can be better accessed and controlled).

We might then assign a low-level state to a mental-state type based not on its present properties, but on either:

> 1) the kind of high-level state that it used to be before automatisation, and
>
> 2) the kind of high-level state that it could be after articulation.

Summary:

It's often attractive to distinguish high-level and low-level versions of some familiar mental-state type.

There is a general prima facie worry about whether such moves are consistent.

A 'primary response' shows a particular move to be consistent after all, given particular definitions of the levels and of the mental-state.

A 'secondary response' accepts that the move is not strictly consistent, but rescues it by
  1. Using a new definition
  2. Relativising the states to a sub-system
  3. Appealing to level-crossing tokens

References:

Bargh, J. A. (1994). 'The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition.' In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition,* Hillsdale, NJ: Erlbaum: 1-40.

Drayson, Z. (2012). 'The uses and abuses of the personal/subpersonal distinction.' Philosophical Perspectives 26: ??

Gallagher, Shaun. 2007. 'Simulation Trouble.' Social Neuroscience 2:3, 353 - 365.

Gallagher, Shaun. 2015. 'The new hybrids: Continuing debates on social perception.' Consciousness and Cognition 36 (2015) 452–465.

Gallagher, S., & Zahavi, D. (2008). 'The (in)visibility of others: A reply to Herschbach.' *Philosophical Explorations* 11(3): 237–244.

Gallese, V., and Goldman, A. (1998). 'Mirror neurons and the simulation theory of mind-reading.' Trends in Cognitive Sciences, 2, 493-501.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). 'Are "implicit" attitudes unconscious?' *Consciousness and cognition* 15: 485-499.

Gendler, T. S. (2008a). 'Alief and Belief'. *Journal of Philosophy* 105(10): 634–63.

Gendler, T. S. (2008b). 'Alief in Action (and Reaction)'. *Mind and Language* 23(5): 552–85.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.* Oxford: Oxford University Press.

Goldman, A. (2008). 'Hurley on Simulation.' *Philosophy and Phenomenological Research* 77 (3): 775-788.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. Cognition, 94, 193-213.

Greenwald, A. G., & Banaji, M. R. (1995). 'Implicit social cognition: Attitudes, self-esteem, and stereotypes.' *Psychological Review* 102: 4–27.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2013). 'Awareness of implicit attitudes.' *Journal of Experimental Psychology* ??

Helmholtz, H. von (1860/1962) Handbuch der physiologischen optik, vol. 3, ed. & trans. J. P. C. Southall. Dover. (Original work published in 1860; Dover English edition in 1962).

Herschbach, M. (2008). 'Folk psychological and phenomenological accounts of social perception.' *Philosophical Explorations* 11(3): 223–236.

Holroyd, J. (2014). 'Implicit bias, awareness and imperfect cognitions.' *Consciousness and Cognition* 33: 511–523.

Holroyd, J. (2016). 'What do we want from a model of implicit cognition?' *Proceedings of the Aristotelian Society* 66 (2): ??

Hurley, S. (2008). 'Understanding Simulation.' Philosophy and Phenomenological Research 77 (3): ??

Jacob, P. (2002). The scope and limits of mental simulation. In J. Dokic & J. Proust (Eds.), Simulation and knowledge of action (pp. 87-109). Amsterdam: John Benjamins.

Jeannerod, M., & Pacherie, E. (2004). 'Agency, simulation and self-identification.' *Mind and Language* 19 (2): 113–146.

Levy, N. (2015). 'Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements'. *Nous* 49 (4): 800–23.

Machery, E. (2016). 'De-Freuding Implicit Attitudes'. In ??. Brownstein and J. Saul (Eds.) ??: 104–29.

Mandelbaum, E. (2015). 'Attitude, Inference, Association: On the Propositional Structure of Implicit Bias.' *Nous* ??

Michael, J., Christensen, W., & Overgaard, S. (2014). 'Mindreading as social expertise.' *Synthese* 191 (5): 817–840.

Saul, J. (2013). 'Unconscious Influences and Women in Philosophy.' In F. Jenkins and K. Hutchison (Eds.), *Women in Philosophy: What Needs to Change?*, Oxford: Oxford University Press: 39–60.

Saxe, R. (2005). 'Against simulation: The argument from error.' *Trends in Cognitive Science* 9: 174–179.

Saxe, R. (2009). 'The neural evidence for simulation is weaker than I think you think it is.' *Philosophical Studies* 144: 447–456.

Slors, Marc. 2010 'Neural resonance: Between implicit simulation and social perception.' Phenom Cogn Sci (2010) 9:437–458.

Stich, S. (1978). 'Beliefs and Subdoxastic States.' *Philosophy of Science* 45 (4): 499-518.

Zahavi, D. (2008). 'Simulation, projection and empathy.' *Consciousness and Cognition* 17: 514–522.