to refute physicalism and that interactionism is implausible, the only reasonable option left for him seems to be epiphenomenalism. However, some philosophers hold that the knowledge argument is not consistent with epiphenomenalism. Epiphenomenalism claims that *qualia are causally inefficacious in the physical world. Ironically, this claim appears to contradict the Mary scenario, for if qualia really are causally inefficacious in the physical world, then surely she does not come to know anything by having colour qualia upon her release. Therefore, according to this objection, one cannot consistently accept both the knowledge argument and epiphenomenalism at the same time. This objection does not show exactly which premise of the knowledge argument is false, but it does show—if it shows anything—that there must be something wrong with the argument. This objection is obviously based on a version of the causal theory of knowledge, which itself is a matter of controversy.

Churchland (1989) provides an objection to the knowledge argument in the same vein. According to him, there must be something wrong with the knowledge argument because if the argument successfully refuted physicalism it would equally successfully refute some versions of dualism as well. Suppose, for example, that substance dualism is true and that in her black-and-white environment Mary learns not only all truths about the physical entities, but also all truths about mental substance. That is, she learns everything about the causal, relational, and functional roles of physical entities as well as of mental substance. However, it still seems obvious that she learns something when she has a colour experience for the first time. Therefore, Churchland concludes, the knowledge argument is unreasonably strong.

As I noted earlier, Jackson no longer endorses the knowledge argument. In his second postscript published in 1998, he declared that he had come to think the knowledge argument failed to refute physicalism. Moreover, in his 2003 paper, he introduced and explained in detail his own objection to the knowledge argument. In constructing his objection he appeals to *representationalism, according to which phenomenal states are representational states. He says that what happens to Mary upon her release is not to learn new non-physical truths, but merely to be in a new kind of representational state. While this position might appear similar to the new mode of presentation response mentioned above, Jackson characterizes it as a version of the ability hypothesis. For, unlike many proponents of the new mode of presentation response, he rejects the idea that Mary acquires any propositional knowledge, whether it is old or new, upon her release. Mary merely comes to be in a new representational state without acquiring or reacquiring any knowledge. Mary acquires instead, according to Jackson, abilities to recognize, imagine, and remember the new representational state.

Along with the conceivability argument and the *explanatory gap argument, the knowledge argument is regarded as one of the greatest objections to physicalism. While there are a number of strong arguments for physicalism, any version of physicalism that is vulnerable to the knowledge argument is inadequate.

Many of the papers referred to in this entry are reprinted in Ludlow et al. (2006).

YUJIN NAGASAWA

Alter, T. (1998). 'A limited defence of the knowledge argument'. *Philosophical Studies,* 90.

Bigelow, J. and Pargetter, R. (1990). 'Acquaintance with qualia'. *Theoria,* 61.

Churchland, P. (1989). 'Knowing qualia: a reply to Jackson'. In *A Neurocomputational Perspective*.

Conee, E. (1994). 'Phenomenal knowledge'. *Australasian Journal of Philosophy,* 72.

Dennett, D. C. (1991). *Consciousness Explained*.

Foss, J. (1989). 'On the logic of what it is like to be a conscious subject'. *Australasian Journal of Philosophy,* 67.

Horgan, T. (1984). 'Jackson on physical information and qualia'. *Philosophical Quarterly,* 34.

Jackson, F. (1982). 'Epiphenomenal qualia'. *Philosophical Quarterly,* 32

—— (1986). 'What Mary didn't know', *Journal of Philosophy,* 83.

—— (2003). 'Mind and illusion'. In O'Hear, A. (ed.) *Minds and Persons*.

Lewis, D. (1988). 'What experience teaches'. *Proceedings of the Russellian Society (University of Sydney),* 13.

Locke, J. (1689). *An Essay on Human Understanding*.

Ludlow, P., Nagasawa, Y. and Stoljar D. (eds) (2000). *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*.

Meehl, P. E. (1966). 'The complete autocerebroscopist'. In Feyerabend, P. and Maxwell, G. (eds) *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*.

Nemirow, L. (1990). 'Physicalism and the cognitive role of acquaintance'. In Lycan, W. G. (ed.) *Mind and Cognition: A Reader*.

Stoljar, D. (2006). *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*.

knowledge, explicit vs implicit. In the scientific study of mind a distinction is drawn between *explicit knowledge*— knowledge that can be elicited from a subject by suitable inquiry or prompting, can be brought to consciousness, and externally expressed in words—and *implicit knowledge*—knowledge that cannot be elicited, cannot be made directly conscious, and cannot be articulated. Michael Polanyi (1967) argued that we usually 'know more than we can say'. The part we

can articulate is explicitly known; the part we cannot is implicit.

Three things are worth noting about the prevailing distinction. First, as studied today in cognitive psychology, it rests on the ability of a subject to present information in linguistic form, to verbally report the thing known. Since there is nothing intrinsic in the idea of externalization and expression that need restrict it to language, this is needlessly confining. When someone has explicit *memory of an event or process, the thing remembered might be a visual scene, a body movement, a taste, smell, or sound. To communicate body-based or sensory recollections it may be necessary to use non-verbal forms of expression, such as illustrations, musical or vocal expression, dance, gesture, and so on. 'I remember: you perform the step like *this*.' The bodily movement is necessary for the subject herself to both know and communicate the details of the step.

Second, to successfully prompt or elicit information, it may be necessary to give subjects tools or artefacts they normally use when in their normal context. Some people can remember telephone numbers only if they have their phone in hand, or remember the combination to a lock if they turn the dial. Other people need a pen in their hand to recall what they wrote earlier, or need shoes to show how to tie shoelaces. There is nothing intrinsic to the idea of prompting or eliciting knowledge that restricts it to verbal requests in a sterile laboratory environment, or prohibits using tools to express the content of a behaviour-governing rule. Subjects often need artefacts to enact their knowledge.

Third, the range of things licensed as implicitly knowable under the prevailing definition is enormous. Things like implicit grammars, implicit rules of inference, implicit memories, implicit knowledge of physical principles such as the speed of sound or the rigidity of objects, implicit knowledge of environmental regularities, even implicit knowledge of the distance between one's ears, are all, in principle, objects of knowledge because each might be implicit or built into a process model. This would not be so problematic if there were a settled theory explaining how knowledge may be 'in' a system. (Kirsh 2006). But there is not. This is a concern because knowledge attributions in science are meant to designate causal states. So a deeper theory of how implicit knowledge is represented or incorporated in a system is required to fully justify claims that a subject 'really' has implicit knowledge. (cf. Dienes and Perner 1999).

1. The basic idea of implicit knowledge

Before cognitive psychologists and neuroscientists developed special methods for studying implicit knowledge, theorists like Polanyi (1967) and Noam Chomsky (1965) had already discussed the importance of *tacit knowledge*. When Polanyi spoke of knowing 'more than we can tell' he was talking about how practical know-how, or procedural knowledge, is tied to our context of work, and resists articulation and *codification*. Our practical knowledge is often highly *situated*, to use a more recent term, and so it is something we frequently are not aware that we know, and cannot tell anyone about.

For instance, the visuo-motor-tactile programs that control how we flip an egg 'over easy' are causal programs; they are procedures that rely on registering subtle details of a situation that we are often not explicitly aware of and usually cannot describe. We can show someone how to flip an egg, possibly tell them about certain explicit factors to watch out for; but there are other, more tactile features relating to the feel of the spatula and egg that practice has taught us to monitor *automatically and unconsciously. We cannot describe them because we are unaware of the highly contextualized 'micro-features' we are attending to. Even if we explicitly know what those contextualized features are we cannot codify them in rules, or even point them out to others because the things to be shown may be tactile, which are not readily communicable, or they are features that only someone simultaneously flipping an egg can identify, and only then if the listener has the prior skills to register those micro-features. For example, a wine expert may prefer one wine to another for reasons he cannot explain. He does not know all the gustatory and olfactory features that go into his classification. Explanations he does give invariably contain words, such as 'round tannins', that non-experts lack the training to understand. Even for experts, the shared vocabulary falls far short of the features that causally affect judgement. Polanyi believed that many of the component elements of expertise are unconscious, non-communicable, and tacit.

Chomsky (1965) also argued for tacit or implicit knowledge, this time for implicit knowledge of linguistic structure and generative grammar. On his view, anyone who knows her mother tongue must, in a sense, know the syntax of her language. If she were unschooled in grammar, or her culture never defined a grammar for her language, she has none of the technical concepts such as noun, subject, verb, and adjectival phrase that figure in the rules of generative grammar. So she cannot state those rules or recognize them if stated by someone else. Hence, she does not explicitly know her grammar.

Nor can she be conscious of those rules when they are operative since she does not have the conceptual repertoire to form thoughts about them. Chomsky thought they were in a modular subsystem inaccessible to conscious probing. Consequently, if she knows her grammar at all she knows it implicitly.

Despite differences in the types of knowledge that Chomsky and Polanyi considered, both maintained that tacit or implicit knowledge is real: it is causally active, it drives behaviour, it is learned, and it is encoded somewhere in the mind–brain in informational states, structures, or processes. Those informational states figure in mechanistic explanations of language production and recognition, or of skilled workplace performance, regardless of what the underlying mechanism is: rule-based system, symbolic constraint system, neural network, or something else. Neither Chomsky nor Polanyi, however, thought it their job to say how tacit knowledge is actually realized in cognitive systems.

We expect cognitive psychologists to provide theories explaining how different types of implicit knowledge are embodied in cognitive or neural systems. What are the mechanisms by which this or that type of implicit knowledge is able to unconsciously influence thought or behaviour? What is the route by which it enters the cognitive system? To probe for such states experimentalists have developed methods for detecting the effect of knowledge without informing a subject that they are interested in that knowledge.

For instance, to test *implicit memory* a subject may be given a list of words and asked to alphabetize them. The experimenter gives no hint that she is interested in the subject's memory for the words on the list, so the subject has no reason to form the intention to memorize the words. Later, the subject is shown a new list consisting of three kinds of words: those drawn from the original list, words not on the list, and pseudo-words—letter sequences that could be words but are not (e.g. bluck). Each word or pseudo-word is shown for 50 ms or a bit less, the normative time subjects take to recognize a word correctly 50% of the time. The subject's task is to state whether the stimulus word is a real word or non-word. It has been found that words on the original list are correctly recognized as words more often than non-list words and both are recognized as words more often than pseudo-words are recognized as non-words. This shows that subjects have some sort of memory for the words on the original list, despite their not trying to remember the list words, and despite not realizing that list words are being tested for. The list words are said to be **primed* because they seem ready to surface faster. Importantly, if the test stimuli are different in appearance to the list stimuli (in font, size, or colour) the effect of priming greatly decreases. Some psychologists see this as evidence that there are two kinds of memory system based on different brain systems (Squire 1992). Others see this as showing that priming is an early stage of processing, and that explicit tasks require stimuli to be more deeply processed (Craik and Lockhart 1972).

Other examples of implicit knowledge discussed in the psychological and neuropsychological literature include, among many others, *blindsight and implicit *learning. In blindsight, patients who have lost part or all of their visual field as a result of a stroke or injury to their visual cortex can often tell whether a visual stimulus is present (though not with great reliability) despite reporting, quite convincingly, that they can see nothing. (Weiskrantz 1986). Blindsight is a form of perception where the subject has no explicit awareness of the visual stimulus but can show by other means that they know something about the stimulus. Whereas normal perception yields explicit knowledge, blindsight yields implicit knowledge, or something close to it, since probing regularly elicits correct answers without the awareness or confidence that comes with normal perception: 'How can I tell you if I don't see it?' The presence of blindsight shows that something is getting in somewhere in those subjects' visual system, but not in a form, or to a processing location, where it can have its full range of normal effects. It is not brought to conscious mind.

In implicit learning experiments subjects are trained to classify items as either in or out of a category. In a famous set of experiments Reber (1989) showed subjects sequences of letters like *aaba, abaa, bba* that were either generated by an *artificial grammar (a Reber grammar), or randomly. After being trained on a set of exemplars, subjects were shown additional sequences and told whether their own classifications were correct or incorrect. They then had to predict whether new sequences were in or out of the language. If subjects reported trying to conjecture the rule governing legal sequences, and they used that rule successfully in their answers, then they had explicit knowledge of the grammar. If they could not report the rule, either because they did not use one, or were unaware they used one, or their answers were inconsistent with their stated rule, then the basis for their category judgement could not have been explicit knowledge of a rule. They were assumed to have implicit knowledge of a categorizing principle, however, because they categorized in a self-consistent manner.

The final type of implicit knowledge to be mentioned is one that further extends the range of implicit knowledge. David Marr (1983) in his influential account of visual processing discussed the importance of posing visual information processing problems as computational problems: a level of analysis where theorists

study the assumptions about the visual world that must be built into human or animal visual systems. He asked: What must particular modules of the visual system implicitly know about the visual world if they are to work correctly? For example, to extract three-dimensional shape from the sequence of two-dimensional retinal images made by a moving object, Marr suggested that the system must assume that objects are rigid and piecewise smooth. He then went on to suggest various algorithms and representations that might operate in a visual subsystem based on those assumptions.

Because Marr did not suppose that these assumptions are explicitly represented anywhere in the creature, it is hard to understand in what sense the creature (or visual subsystem) has knowledge, albeit implicit. Is it causal? One might argue that these assumptions are better understood as *success conditions:* if a moving object meets these conditions then algorithms that presuppose their truth will generate the right shape. If the object does not meet the conditions then algorithms presupposing it will not terminate, or the subject will see an *illusion.

Rigidity and continuity are presumably not learned by the visual system; they are the outcome of natural selection sifting through algorithms for the ones that work best. The same might be said for Chomsky's universal grammar. They set constraints on all viable generative grammars. Yet Chomsky maintained that universal grammar as well as particular grammars are causal. They shape language learning. Why not assume a similar causal role for rigidity? This makes it more important than ever to explain how implicit knowledge might be represented, instantiated, or embodied in cognitive systems.

Given the variety of implicit knowledge, it is likely there are major differences in the way such information states are encoded or embodied in cognitive systems. By definition, all implicit forms resist linguistic processing, but there are many possible reasons for this. It has been speculated that knowledge is implicit because it is stored in parts of the cognitive or neural system that do not directly communicate with linguistic parts, hence the thing known is not articulable (the modularity of cognitive sub-systems). Alternatively, some contentful states might require too much processing to be converted into words in reasonable time (computationally too distant), or because content is encoded initially in too shallow a manner and hence overly dependent on interaction with other (currently inaccessible) representations of knowledge to become explicit. These are just a few of the process model explanations that would show how knowledge that cannot be made conscious can nonetheless causally affect thought and behaviour.

Despite recent empirical advances, we are still in the early days of understanding the causal pathways leading to consciousness and behaviour. We can be certain that our conception of implicit and explicit knowledge will change as new process models and theories are proposed, and scientists shift their understanding of what it means to say that someone knows something implicitly. For instance, it is a significant defect of current process models that they do not fully accommodate the importance of non-verbal awareness and expression. That means that the concept of explicit knowledge in use today is so narrow that it forces us to call some knowledge states implicit when a more multimodal notion of consciousness, one that admits non-verbal imagery and artefact use, would warrant calling them explicit.

Similarly, the concept of implicit knowledge is today so broad that it is unclear whether we could ever have process models that reveal how all the different types of implicit knowledge play a causal role in affecting thought, talk, and action. For instance, we assume that a person will come to know the implications of their beliefs, if given time to reflect on them. Are those implications therefore implicitly known before reflection but explicitly known after reflection? That would be odd, because other types of implicit knowledge are never explicitly knowable, regardless of reflection. Similarly, humans are assumed to share a vast realm of implicit common knowledge with their cultural peers. Yet it is doubtful whether all members of a culture share this common ground equally. At the cultural level we say they know it implicitly, at a more process level they do not. That means that the concept of implicit knowledge in use today is so broad and heterogeneous that the term will be negotiated and renegotiated as new process theories re-characterize how implicit knowledge can be causally active.

2. The connection to representation

One promising way of lending rigour to the distinction, even before future negotiations, is to tie it with the notion of explicit and implicit *representation. On virtually every account, explicit knowledge is connected with thought. Although knowledge and thought are different in kind—knowledge is a dispositional state and thought an occurrent process—thought is the way that explicit knowledge typically manifests itself. This means that if someone explicitly knows something then she *can* bring the thing known 'before' mind. She can 'grasp' the content of the known thing. This raises

the provocative idea that something is known explicitly by an agent, only if she can represent it, and in a form that is 'immediately' graspable, presumably by the conscious mind. To represent something in a form that is immediately graspable is to *represent* it explicitly (Kirsh 1990).

Viewing things in this light partially resolves several issues. First, it explains the historical bias for verbalizing knowledge and lets us get beyond it. The classical justification, were it ever to be given, would go like this: knowledge is explicit for someone if she can bring it to mind as a thought—she can think it; if she can think it she can speak it—assumed because language is the most structured account of content available (see Fodor 1975), and a public language, like English, is universally expressive (see Searle 1970). Sentences in English are, accordingly, explicit representations of what is known. Hence anything known explicitly should be articulable.

This vaguely behaviourist move saves having to identify explicit knowledge with what can be brought to consciousness per se because it associates bringing to mind with being verbalizable. But it is imperfect for reasons that reveal there is a more fundamental notion of explicit representation.

First, language is not perfectly expressive. How can a squiggly curve, for instance, be expressed accurately without gesturing or making a drawing? Demonstratives such as 'this' often take non-linguistic things as completions. For example, when a person hears a sound the only adequate way of identifying where it comes from is usually by pointing. Would anyone doubt the person had explicit knowledge of where the sound was? Their explicit knowledge consists in having an 'active' set of orienting responses, the most easily shared being to point. The same applies to dance movements, sounds, and sights. Words are useless, or of limited use, in trying to expose, even to oneself, what is explicitly known. Some things must be shown, not told. This calls into doubt the *necessity* of encoding explicit knowledge in language, and identifying the content of knowledge with linguistically expressed propositions.

Second, many things presented in language are not immediately graspable, so linguistic expression may not be *sufficient* for explicit knowledge. For example, the sentence, 'Police police police police police' is grammatical and means police who are policed by police themselves police police. Considerable processing must occur before this sentence can be grasped. Indeed, most people cannot readily extract its meaning any more than they can extract the meaning of a complex mathematical formula. This suggests that being encodable in a natural language is not a sufficient condition of being explicit. To be explicitly known the content of thought must be encoded in a form that is immediately graspable according to some prior measure of immediacy.

Kirsh suggested that the degree to which a given representation *R* explicitly encodes information *I*, for a given creature *C*, should be measured by the amount of computation *C* must perform to extract *I*. For instance 'fifth root of 3125' is a less explicit encoding of 5 than the numeral '5'. The creature must compute the fifth root before it can grasp the referent. Hence the information is not on the surface in '3125' but is on '5'. It is implicit, but less so than $\sqrt[17]{762{,}939{,}453{,}125}$.

The value of such an approach is that it ties explicitness to computation in a manner that is not parochially bound up with language. But it also leaves open the need to tie explicitness to the computational resources a creature has. Thus if one creature has memorized exponents of 5 up to 5^{17}, the computation of $\sqrt[17]{762{,}939{,}453{,}125}$ may be a simple retrieval process. Similarly, a creature with a highly parallel computational system, such as human vision or motor control, may be able to process complex structures rapidly when they are visually or motor encoded, but more slowly when linguistically encoded. So content shown visually might be explicit while being more implicit when given linguistically. It also gives a place for learning, since highly practised agents can immediately grasp contents, such as wine tastes, musical structures, concepts and so forth, that would be difficult for the unpractised. They have *automatized or parallelized them.

The upshot is that when explicit knowledge is tied to explicit representation it makes the notion less behavioural and more closely tied to discovering the processing pathways by which information stored or built into a system makes its way to an explicit representation. If no such pathway exists, or if the result of further processing falls short of complete explicitness, the system's knowledge is to some specifiable degree implicit. This rightly emphasizes that knowledge lies on a continuum with fully explicit at one end.

D. KIRSH

Chomsky, N. (1965). *Aspects of the Theory of Syntax*.

Craik, F. I. M. and Lockhart, R. S. (1972). *Levels of Processing: A Framework for Memory Research*.

Davies, M. (2001). 'Knowledge (explicit and implicit): philosophical aspects.' In Smelser, N. J. and Baltes, P. B. (eds) *International Encyclopedia of the Social and Behavioral Sciences*.

Dienes, Z. and Perner, J. (1999). 'A theory of implicit and explicit knowledge'. *Behavioral and Brain Sciences, 22*.

Fodor, J. (1975) *The Language of Thought*.

Kirsh, D. (1990). 'When is information explicitly represented?' In Hanson, P. (ed.) *Information, Language, and Cognition*.

—— (2006) 'Implicit and explicit representation'. In Nadel, L. (ed.) *Encyclopedia of Cognitive Science*.

Marr, D. (1983) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*.

Polanyi, M. (1967). *The Tacit Dimension*.

Reber, A. S. (1989). 'Implicit learning and tacit knowledge'. *Journal of Experimental Psychology: General,* 118.

Searle, J. (1970). *Speech Acts*.

Squire, L. R. (1992). 'Declarative and nondeclarative memory: multiple brain systems supporting learning and memory'. *Journal of Cognitive Neuroscience,* 99.

Weiskrantz, L. (1986). *Blindsight: A Case Study and its Implications*.