BEHAVIORAL AND BRAIN SCIENCES (1999) **22,** 735–808
Printed in the United States of America

A theory of implicit and explicit knowledge

Zoltan Dienes
Experimental Psychology, University of Sussex, Brighton, Sussex BN1 9QG, England
dienes@epunix.susx.ac.uk
www.bids.susx.ac.uk/faculty/ep/dienes.htm

Josef Perner
Institut fuer Psychologie, Universitaet Salzburg, A-5020 Salzburg, Austria
josef.perner@sbg.ac.at **www.sbg.ac.at/psy/people/perner_ e.htm**

Abstract: The implicit-explicit distinction is applied to knowledge representations. Knowledge is taken to be an attitude towards a proposition which is true. The proposition itself predicates a property to some entity. A number of ways in which knowledge can be implicit or explicit emerge. If a higher aspect is known explicitly then each lower one must also be known explicitly. This partial hierarchy reduces the number of ways in which knowledge can be explicit. In the most important type of implicit knowledge, representations merely reflect the property of objects or events without predicating them of any particular entity. The clearest cases of explicit knowledge of a fact are representations of one's own attitude of knowing that fact. These distinctions are discussed in their relationship to similar distinctions such as procedural-declarative, conscious-unconscious, verbalizable-nonverbalizable, direct-indirect tests, and automatic-voluntary control. This is followed by an outline of how these distinctions can be used to integrate and relate the often divergent uses of the implicit-explicit distinction in different research areas. We illustrate this for visual perception, memory, cognitive development, and artificial grammar learning.

Keywords: artificial grammar learning; automaticity; cognitive development, consciousness; implicit knowledge; memory; visual perception

1. Introduction: Objectives

The objective of this target article is to provide an analysis of the distinction between implicit and explicit knowledge in terms of the semantic and functional properties of mental representation. In particular this analysis attempts to:

(1) Create a common terminology for systematically relating the somewhat different uses of the implicit-explicit distinction in different research areas; in particular, learning, memory, visual perception, and cognitive development;

(2) Clarify and generate predictions about the nature of implicit knowledge in different domains;

(3) Clarify why the distinction has traditionally been brought into close contact with notions such as consciousness, verbalizability, voluntariness-automaticity, and so on;

(4) Justify why different empirical criteria (e.g., subjective threshold, objective threshold, direct-indirect tests) are used to identify implicit-explicit knowledge;

(5) Justify the use of the implicit-explicit terminology by observing the ordinary language meaning of "implicit" and "explicit."

Our basic strategy for meeting these objectives is to analyse knowledge as a propositional attitude according to the representational theory of mind (RTM; Field 1978; Fodor 1978). Roughly speaking, if I know a fact (e.g., the animal in front of me is a cat) then, according to RTM, I have a representation of that fact and the internal, functional use of this representation constitutes it as knowledge of mine (rather than a desire of mine, etc.). Knowledge can vary depending on what is represented (made explicit) and which aspects remain implicit in the functional use of rep-

ZOLTAN DIENES is a Reader in Experimental Psychology at the University of Sussex. He is the author of more than 30 scientific publications, mainly in the area of implicit learning, including the co-authored book (with Dianne Berry) *Implicit learning: Theoretical and empirical issues,* and papers in *Nature, Journal of Experimental Psychology,* and *Cognitive Science.* He also won an award for the best designed experiment to test Sheldrake's theory of Morphic Resonance.

JOSEF PERNER is Professor of Psychology at the University of Salzburg. He wrote *Understanding the representational mind* (MIT Press, 1991) and more than 75 articles on cognitive development, in particular the development of a theory of mind, metalinguistic awareness and executive control, on the use of simulation in cognitive development and decision theory, and on mutual knowledge. He was elected president of the European Society for Philosophy and Psychology for 1999–2000.

 0140-525X/99 $12.50

resentations. This application of the implicit-explicit distinction has several advantages.

The main advantage of our analysis is that it provides a common ground for the use of the implicit-explicit distinction in different fields of investigation. Consider Schacter's (1987) influential definition of the implicit-explicit memory distinction: "Implicit memory is revealed when previous experiences facilitate performance on a task that does not require conscious or intentional recollection of those experiences; explicit memory is revealed when performance on a task requires conscious recollection of previous experiences." This definition may capture the phenomenal experience of implicit and explicit memory very well, but it leaves open how the definition is to apply to implicit and explicit knowledge in other fields. For example, Karmiloff-Smith (1986; 1992) has argued that there are several steps of explicitness before consciousness is reached. Identifying being explicit with being conscious gives us no understanding of why Karmiloff-Smith's lower forms of explicitness have anything to do with this distinction. In other words, although it has been suggested that the implicit-explicit dichotomy should be broken up into a series of explicitness levels, our analysis is needed to explain just what it is that becomes more explicit as one ascends levels, and to relate levels in one research area to different subdivisions of explicitness in other areas.

Existing problems of this kind with the implicit-explicit distinction are many. In research on memory and subliminal perception, explicitness has been linked to performance on direct versus indirect tests (Reingold & Merikle 1993; Richardson-Klavehn & Bjork 1988) because direct test performance seems to require conscious awareness. The interesting question left open, however, is why direct tests require consciousness. Or, in visual perception, it is found that touching an object is based on unconscious, implicit information, whereas pointing to the object requires conscious, explicit information that is subject to visual illusions (e.g., Bridgeman 1991; Milner & Goodale 1995; Rossetti 1997). Why? More directly, what are the representational requirements for conscious awareness? What is the relation between knowledge over which we have voluntary control and knowledge of which we are aware? Why can we sometimes control in limited ways knowledge of which we are not aware (Dienes et al. 1995)? Can predictions be made for the conditions under which knowledge will be represented implicitly? With our analysis of the implicit-explicit distinction, we are able to give answers to some of these questions.

Another advantage of our analysis is that it is grounded in the ordinary use of the terms "implicit" and "explicit" (e.g., "They didn't say so explicitly; it was left implicit"), whereas traditional definitions have depended on further related distinctions. Schacter (1987, p. 501) defined implicit memory by its lack of conscious or intentional recollection, and Reber (1993, p. 5) defined implicit learning as "the acquisition of knowledge that takes place largely independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired." These definitions of implicit memory/learning raise the question of why the terms implicit/explicit are used at all. Why not call explicit memory or learning directly by their name, that is, conscious memory or conscious learning (cf. Reingold & Merikle 1993, p. 42)? Moreover, when using technical terms with an existing ordinary meaning, it seems to us, we should adhere to that existing meaning as far as possible and not impose some arbitrary "operational definition," or else we make it difficult for the scientific community to share the same meaning, because the natural meaning is likely to keep intruding. (Who still adheres – or ever adhered – to the operational definition of intelligence as that which the WAIS [Wechsler Adult Intelligence Scale] measures?) So it is not an unimportant feature of our use of the implicit-explicit distinction that it attempts to stay true to its natural meaning, which we believe was the unarticulated reason for introducing the distinction in the first place, and what partially motivated its acceptance and continued use.

We ordinarily say that a fact is conveyed explicitly if it is expressed by the standard meaning of the words used. If something is conveyed but not explicitly, then we say it has been conveyed implicitly. We can discern two main sources of implicitness. One is the contextual function/use of what has been said explicitly. A prime case is *presuppositions.* To use a famous example, the statement, "The present king of France is bald," presupposes that there is a present king of France. It does not express this fact explicitly because the function of the sentence (when uttered as an assertion) is to differentiate the present king of France being bald from his not being bald. For that reason the speaker of this sentence can claim that he *did not* (explicitly) say that there was a king of France. Yet the presupposition does commit him to there being a king of France, or else his assertion of the king being bald becomes insincere. So in this sense he did (and thus we say: "implicitly") convey that there is a king of France.

The other source of implicitness lies in the conceptual structure of the explicitly used words. For example, if one conveys that a person is a *bachelor,* then one conveys that this person is *male* and *unmarried* without making those features explicit. Using "bachelor" commits oneself quite strongly to "male" and "unmarried" lest one shows oneself ignorant of the meaning of the word bachelor in the language spoken. These are not rare cases. Whenever we say that something is an X (e.g., a bird), we implicitly convey that it is also an instance of the superordinate category of X (e.g., an animal) on the same grounds as in the bachelor case.

It is common to both sources of implicitness that the information conveyed implicitly concerns *supporting facts* that are *necessary* for the explicit part to have the meaning it has. The implicitly conveyed fact that *there is a king of France* is necessary for the explicitly expressed information that *he is bald* to have its normal, sincere meaning. Similarly, that someone is male and unmarried is a necessary supporting fact for the explicitly conveyed fact that he is a bachelor.

In our analysis the distinction is between which parts of the knowledge are explicitly represented and which parts are implicit in either the functional role or the conceptual structure of the explicit representations. A fact is explicitly represented if there is an expression (mental or otherwise) whose meaning is just that fact; in other words, if there is an internal state whose function is to indicate that fact.[1] Supporting facts that are not explicitly represented but must hold for the explicitly known fact to be known are *implicitly represented.*

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

2. The representational theory of knowledge

2.1. Implicitness arising from functional role

Mental concepts such as knowledge are standardly analysed as propositional attitudes (Russell 1919). The sentence "I know that this is a cat" consists of a person (I), a proposition (this is a cat), and an attitude relation between person and proposition (knowing). The representational theory of mind (Field 1978; Fodor 1978) is concerned with how such an attitude can be implemented in our mind. The suggestion is that the proposition is represented and the attitude results from how that representation is used by the person (functional role). The representation "this is a cat" constitutes knowledge if it is put in what philosophers would call a "*knowledge box*" or cognitive scientists would call a *database.* The representation is used as a reflection of the state of the world and not as it would be, for example, if it were in a *goal* box, as a typically nonexisting but desirable state of the world.

In this view, we can say that the content of the knowledge is explicit because it is represented by the relevant representational distinctions (by analogy with explicit verbal communication). That is, there is an internal state whose function is to indicate the content of the knowledge. In contrast, the fact that this content functions as knowledge is left implicit in its functional role[2] (as implicitly conveyed information is communicated by the functional necessities created by the explicit part). The fact that it is I myself who hold this knowledge is not explicitly represented; it is implicit in the fact that I do hold that knowledge. We accordingly have three main types of explicit knowledge, depending on which of the three constituents of the propositional attitude is represented explicitly:

(1) explicit content but implicit attitude and implicit holder (self) of the attitude;

(2) explicit content and attitude but implicit holder of attitude; or

(3) explicit content, attitude, and self.

This large picture has to be refined in at least three ways. In the first place, the same shift from implicit to explicit also applies within each constituent, complicating the picture somewhat. In addition, arguments are needed as to why only the above combinations occur and not all the other logically possible ones (e.g., an explicit representation of self but implicit attitude and content). We start by discussing the refinements required for the first type of each of the three constituents of propositional attitudes.

2.1.1. Content. The content of a propositional attitude, like knowledge, is what the attitude is about. In the example of the cat that I see in front of me, I know that it is a cat. The representation of the content of this knowledge as "this is a cat" identifies (1) a *particular individual* (i.e., the animal in front of me), (2) a *property* (or natural kind: catness), and (3) it *predicates* this property of the particular individual. For a more succinct and more general way of expressing these aspects we use predicate calculus notation, where F, G, . . . denote properties, a, b, . . . denote particular individuals, and the syntactic combination of F and b into the formula Fb expresses that F is predicated of b.

Even though this content makes these three elements explicit, however, there are other aspects that remain implicit. For example, I clearly know that the individual is *now* a cat, and that it is a *fact* about the *real* world that it is a cat, not just a cat in some fictional context. That is, (4a), the temporal context of the known state of affairs, and (4b), its factuality, are left implicit.

We have identified four main components of a known fact about which we can ask whether they need to be represented explicitly or can be left implicit:

(1) properties, e.g.: "F," "being a cat."

(2) individuals, e.g.: "b," "particular individual in front of me."

(3) the predication of the property *to* the individual, e.g.: "Fb," "this is a cat."

(4a) temporal context.

(4b) factuality (versus fiction), e.g.: "It is a fact of this world that at time t, Fb," "It is a fact that this is currently a cat."

The question is now whether any of these components can remain implicit and whether they can remain implicit independently of each other or only in certain combinations. We argue that they can only remain implicit in roughly the order in which they are listed above, that is, if an element with a higher number is represented explicitly then every element of a lower number must also be represented explicitly.

As an extreme case in which almost everything is left implicit we consider Strawson's (1959, p. 206) "naming game," in which a person simply calls out the name of a presented object, for example, "cat" or "dog," depending on which kind of animal is presented. In this context, the word "cat" expresses knowledge of the fact that "this (object in front of the person) is a cat" and conveys this information to the initiated listener. We could not say anything less, for example, that it only expresses knowledge of catness, or of the concept of cat. Yet, what are made explicit within the vocabulary of this naming game are only the properties of being-a-cat, being-a-dog, and so on. Consequently, because there is knowledge that it is the particular presented individual that is a cat or dog, that knowledge remains implicit.[3]

Our use of Strawson's naming game only provides an example of the property (cat) being represented explicitly, the individual and predicating the property of this individual remain implicit. The naming game uses the publicly inspectable medium of language, but, when it comes to the question of which aspects can be made explicit independently of other aspects, it becomes an imperfect guide for explicitness of mental representations, as the following shows.

In the naming game, it is also possible to represent individuals explicitly and to leave their properties implicit. This is the case for forced choices between two items, by pointing to the item that has a particular property, for example, which one of two objects – the left or the right – is a cat. In the case of the naming game, one could argue that the *response* must explicitly distinguish the two items (a, b) by pointing right or left, but not the property. The pointing thus conveys the information "This one is a cat" but makes only "this one" explicit and leaves "is a cat" implicit. In the case of the naming game (i.e., the information passing between two communicating parties), this is possible. In the case of the knowledge that a single person must bring to bear, explicitness of the individuals requires explicitness of the attributed property, because the person must be able to

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

go into a cat/no-cat state for each individual in order to decide which is a cat and then to respond correctly. Hence, for knowledge we have the constraint that explicit representation of the individual to which a property is attributed entails explicit representation of that property.

At this point one should be aware that the notion of predicating something of a particular individual need not be restricted to particular objects or persons. It will be applied later in extended form to events and even to causal regularities. Traditional logic does not make this very explicit but Barwise and Perry's (1983) situation semantics offers an elaborate distinction between event types and individual events, in order to capture the capacity of natural language to freely reference particular events, causal regularities, laws, and so on, and then to describe them as having certain properties or as being of a certain type. For example, a particular event (b) was a dance (F) and has the further feature of having had me as a participant (G), and so forth.

Subliminal perception provides an example from psychological research, as discussed in more detail in section 3.2. The suggestion is that under subliminal conditions only the properties of a stimulus (the kind of stimulus) get explicitly represented (e.g., the word "butter"), not the fact that there is a particular stimulus event that is of that kind. This would be enough to influence indirect tests, in which no reference is made to the stimulus event (e.g., naming milk products), by raising the likelihood of responding with the subliminally presented stimulus ("butter" is listed as a milk product more often than without subliminal presentation). The stimulus word is not given as response to a direct test (e.g., Which word did I just flash?) because there is no representation of any word having been flashed. Performance on a direct test can be improved with instructions to guess (Marcel 1993), because this gives leave to treat the direct test like an indirect test, just saying what comes to mind first.

As mentioned earlier, even explicit representation of F being predicated of b ("Fb," or "This is a cat") leaves implicit the fact that Fb is a true proposition, that is, a fact at the present time. Only the representation "Fb is a fact now" represents *the fact that b is F at the present time* completely explicitly. The reason for making these aspects explicit may seem superfluous. In particular, the addition "is a fact" may strike some readers as totally redundant and trivial, so let us dwell briefly on its significance.

Consider a simple mental system that does not represent truth explicitly but just contains a single model of how it perceives the world to be (Perner, 1991, described the young infant as having only this representational power). The model of the world is a type of knowledge box in that any proposition Fb that is in the knowledge box is taken (judged) to be true, on the grounds of being in that box plus the functional role the box plays in the mental economy. There is no possibility of representing propositions that are not true, however, without creating mental havoc, because all propositions in the box are acted on as if they were true (Leslie, 1987, pointed this out in his analysis of pretence). To differentiate true from false propositions, one could represent false propositions in a different functional box, as has been suggested for pretence and for counterfactual reasoning (Currie & Ravenscroft, in press a; Nichols & Stich 1998). In concrete terms this means that a child who is pretending that the banana is a telephone represents "this is a banana (Bb)" in its knowledge box and "this is a telephone (Tb)" in its pretend box. This solution may be adequate for pretend play consisting of switching from a knowledge (serious action) mode into a pretend mode of functioning. Pretend actions are then simply governed by the representations inside the pretend box. This cannot account for the child knowing what it is pretending. To know that, the pretend representations have to be in the knowledge box. This raises the problem of cognitive confusion (representational abuse; Leslie 1987) and the pretend representations have to be quarantined in some sort of "metarepresentational[4] context" (Sperber 1997). Such markers explicitly differentiate within the knowledge box what is to be taken as true from what is not to be taken as true. More generally, for knowing what is and what is not true, the truth value has to be made explicit within the knowledge box, that is, to represent "Fb is a fact" or "Fb is NOT a fact."[5] This distinction is also required for understanding change over time (i.e., to represent that Fb was the case and now Gb is the case; Perner 1991; 1995, Appendix) and to interpret symbolic expressions and representations (e.g., to understand that objects in the world are also in the picture).[6]

The following table gives a summary of the different cases of the possible implicit-explicit combinations of facts that we have discussed so far. We suggest that these are the only realistically possible ones. Table 1 excludes certain permutations of the four components: property, individual, predication, and factuality. For the verbal exclamations in Strawson's naming game, all combinations are possible, but for knowledge only the four cases listed above are possible. For example, predication cannot be known explicitly on its own. It can be explicitly conveyed on its own in the naming game in response to the question "Does b have the property F?" The response "Does/doesn't have it" represents only predication explicitly. Again, a system that can do this must make further internal distinctions; it must distinguish F from not-F to decide whether the presented object "does/doesn't have" that property. Knowledge of the presented individual can remain implicit. This case is accounted for in 2(b) in Table 1.

In the case of factuality we are after the distinction between whether a state of affairs Fb is a fact or fiction. The naming game can only be played with real objects. A system that can meaningfully distinguish between whether the predication of F of b holds in the real world or in a world of fiction must have the representational resources to specify the property and the individual in question and be able to predicate this property of the individual in order to decide whether it holds in reality or only in fiction. Hence, if factuality is known explicitly, then predication, individual, and property must also be known explicitly. Similarly, the time of a fact can only be left implicit for the present. A system that can distinguish between whether the predication of F of b holds now or in the past must have the representational resources to specify the property and the individual in question and be able to predicate this property of the individual in order to decide whether it holds now or has held previously. Hence, if time is known explicitly, then predication, individual, and property must also be known explicitly.

Memory research provides a relevant example. Explicit memory is not only conscious, but, more to the point, is a recollection of the past. For this it must represent past events as having taken place in the past. Only then can systematic answers be given to direct questions about the past.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

Table 1. *Possible combinations of implicit and explicit knowledge of aspects of facts (Factuality stands for factuality and/or time)*

	Represented	
	Explicitly	Implicitly
1.	Property	Individual + predication + factuality
2. (a)	Property + individual	Predication + factuality
(b)	Property + predication	Individual + factuality
3.	Property + individual + predic.	Factuality
4.	Property + ... + factuality	None

If a past event is only represented by its properties (event structure), then it can influence indirect and direct tests alike. Only when the pastness of the event is represented explicitly can performance on a direct test that addresses the pastness directly outshine performance on indirect tests (see Reingold & Merikle's 1993 criterion for explicit memory). So we can see why and how test directness relates to explicitness. In the next section we see how it relates to consciousness.

2.1.2. Attitude. Knowledge is standardly analysed as propositional attitudes. The system knows some fact (e.g., the fact that b is F, or that this is a cat) if it is related in a particular way to the proposition expressing that fact. In the representational theory of mind this is the case if the following conditions hold:

(o) The system has a representation, R, of this fact, and

(i) R is accurate (true),

(ii) R is used by the system as an accurate reflection of reality (i.e., the system must *judge* that b being an F is the case), and

(iii) R has been properly caused (it must not have come about by accident: it must have a respectable causal *origin,* which when made explicit serves to *justify* the claim to knowledge).

Possession, accuracy, judgement, and *causal origin (justification)* are all supporting facts for any representation to constitute knowledge. For example, "Fb is a fact" constitutes knowledge of *the fact that b is F* for a system only if (o) the system has the representation, (i) it is accurate, (ii) it is treated by the system as an accurate reflection of the world (the world is judged to be so), and (iii) it came about in a proper causal (justifiable) way. Hence all four facts are implicit in any knowledge until made explicit.

These four facts define the *attitude* of knowledge. Making them explicit means making the attitude explicit. For that the system has to form the following metarepresentations, where R stands for the representation of the known fact (i.e., R = "Fb is a fact"):

(0) R is possessed by the system.

(1) R accurately reflects the fact that Fb.

(2) R is being taken (judged) as accurately reflecting the fact that Fb.

(3) R was properly caused by its content through a generally reliable process (i.e., it is caused by the fact Fb through the reliable process of visual perception).

In other words, (0) represents that the knowledge content can be entertained by the system, (1) represents the knowledge as a true thought (that is, as a true thought that is being *merely entertained* but *not judged* as being true; see Künne 1995), (2) represents the knowledge as a belief, and (3) represents the knowledge as causally justified thought.

Only if the system can entertain R as a representation that it possesses can it represent what further properties – e.g., (1), (2), and (3) – this representation might have. But the three further metarepresentations can be explicit independently of each other. Truth does not imply having been properly caused nor being taken for true; being taken for true does not imply either being true or being properly caused; and having been properly caused does not imply being taken for true or being necessarily true because, although generally reliable, even such a process can on occasion fail.[7] Note that some dependencies emerge if one represents that it is the same rational agent (e.g., oneself) who represents R as accurate and who represents R as being taken to be true.

If (0)–(3) hold, then the system represents its *attitude of knowing* explicitly, that is: There is knowledge of the fact that Fb. What this does not make explicit is the holder of this attitude, that is, the self. The fact that it is oneself who holds the attitude is implicit in the act of knowing. To make it explicit, the system has to represent itself as the holder of the attitude: "I know that Fb is a fact."[8,9]

Other attitudes may be held towards a piece of knowledge, such as, I *guess* that Fb is a fact. Making any attitude explicit always requires (0) to hold, plus additional representations, depending on the attitude.

2.1.3. Relating explicitness of content, attitude, and self. It is evident that explicit representation of self as holder of an attitude (e.g., "I know . . . ") contains an explicit representation of the attitude ("know"). The interesting question concerns the degree to which explicit representation of knowing requires explicit representation of the content (e.g., this is a cat). That is: Is it possible to explicitly represent "I know" or "it is known" and leave implicit the fact that *this is a cat (Fb)*? In a variation of the naming game, an expression like "I know" can implicitly convey that the knowledge is of the fact that Fb. However, inside a (rational) agent this explicit reflection on knowledge implies explicit factuality of the known; one must be able to judge the factuality of the known fact before coming to the conclusion that one knows that fact. Because explicit factuality implies explicitness of predication, individuals, and properties, we can conclude that explicit representation of self or attitude implies explicit representation of the content.

The dependencies we have discussed are summarised in Figure 1. If an aspect at a higher level is represented ex-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

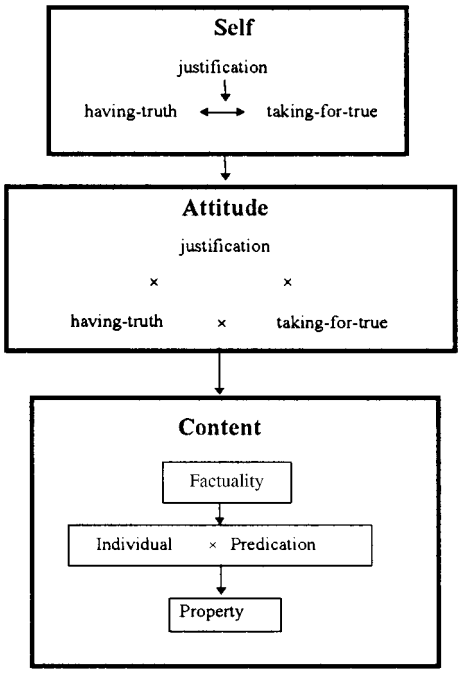

Figure 1. Constraints on explicitness. An arrow denotes that the explicitness of the item from which the arrow emanates entails the explicitness of the item to which the arrow points. An "×" denotes that the explicitness of the two terms can be varied freely.

plicitly (at the origin of an arrow) then – according to our analysis – all aspects at a lower level (at the end of the arrow) must also be explicitly represented.

On the basis of this partial hierarchy we will later speak conveniently of knowledge that is "fully explicit" when all aspects are explicitly represented, "attitude-explicit" when everything up to the attitude is explicit, and "content-explicit" if all the aspects of content are represented explicitly. "Attitude-implicit" will indicate that attitude and all higher aspects in the hierarchy are left implicit, and so on for the other aspects. It is also often convenient to differentiate between different levels within content: "fact-explicit" (equivalent to "content-explicit") when all aspects of content are explicit, "predication-explicit" when predication, individuals, and property are made explicit (for simplicity's sake we ignore the possibility of case 2b in Table 1), and of "completely implicit" if only properties remain explicit.

On an important cautionary note, one must point out that these hierarchical constraints only hold for a single representation. That is, a single representation cannot make something explicit at the higher level and still represent aspects at a lower level implicitly. This does not preclude the possibility of two independent representations, one making something explicit at the higher level and the other representing something at the lower level implicitly. For example:

(a) "I know that there is some fact involving F"
(i.e., explicitly representing attitude and factuality).

(b) "F" (i.e., implicitly representing predication of F to b).

This is possible, but the point is that (a) does not implicitly represent the fact that Fb. Rather, it explicitly represents the knowledge that there is something concerning the property F. In that case, there is no implicit knowledge of Fb being a fact. That this is not implicit in (a) can be seen from the fact that Fb is not a supporting fact of (a), that is, one can know that there was something about F without the fact that Fb.

2.2. Implicitness owing to conceptual structure

This kind of implicitness (*structure implicitness*) typically arises when the system represents (has a concept for) properties that can be defined as compounds of more basic properties, such as the property of being a bachelor has the components of being male and unmarried. So if one explicitly states that a person is a bachelor, then one implicitly conveys that he is also unmarried, because being unmarried is a necessary, supportive fact for being a bachelor. Similarly, one can explicitly know that someone is a bachelor, but not explicitly know that he is not married. However, as not being married is a necessary fact for being a bachelor, this fact is known implicitly. In this example, the structure of the component properties (male, unmarried, etc.) remains implicit in the explicit representation of the compound property (being a bachelor): a case of "property-structure implicitness." Roberts and MacLeod (1995) argued that concepts acquired incidentally and nonstrategically may have nondecomposable atomic representations in which the property structure is represented implicitly in our terminology.

2.3. Summary

We have so far developed a rich structure for describing different ways some knowledge can be implicit within the use of some other explicitly represented knowledge. That is, knowledge with explicit representations of part of its content can contain other parts of its content, the attitude, and self as holder of the attitude implicitly. Also, explicit knowledge can be a representation of compounds (typically: compound properties) that leaves the structure of its components implicit. We now explore how our analysis unifies the different distinctions that have traditionally been used in connection with the implicit-explicit distinction.

3. Related distinctions and test criteria

The previous section showed that knowledge can differ in how many of its functional and conceptual aspects are represented explicitly. This puts us into a position to show that the various distinctions associated with the implicit-explicit distinction differ in the amount of explicit representation required. We start with consciousness, as it has been used most prominently to define explicit knowledge (in memory, Schacter 1987; in learning of rules, Reber 1989). We will show that under a common understanding of "conscious," knowledge counts as conscious only if its content, the attitude of knowing, and the holder of that attitude (self) can be represented explicitly. Hence, conscious knowledge is, indeed, prototypically explicit.

Consciousness has often been related to (even defined in terms of) verbalisability (e.g., Dennett 1978). The ability to address the content of one's knowledge verbally (direct tests) has often been used to test conscious and explicit knowledge. This makes sense in our analysis, because ver-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

bal reference requires very explicit representation of content. Furthermore, a close relative of verbally expressible knowledge, "declarative" knowledge, has often been put in opposition to "procedural knowledge." Although this opposition confounds several independent dimensions (procedural-inert, declarative-nondeclarative, and accessible-inaccessible), we can explain why these groupings appear natural and why they can be tied to the implicit-explicit distinction. Finally, the ability to exert voluntary control, in contrast to automatic action, has been associated with explicit, conscious knowledge. This link is justified, because voluntary control requires explicit representation of one's attitude, which conforms to the requirement for conscious awareness, whereas automatic action can be sustained by procedural know-how.

3.1. Consciousness

We use "consciousness" (some philosophers might find the term "conscious awareness" more appropriate[10]) here as (we think) most people use it; one's knowledge is available to oneself and it is not necessary to prove its existence to one's own surprise through behavioural evidence. This is certainly the meaning of the conscious-unconscious distinction in cognitive psychology, as we will see from the many research examples in the next section. For example, implicit unconscious memory occurs where I appear to have no knowledge (memory) of a past event but can be shown by behavioural evidence in an indirect test to have some (implicit) knowledge of that event.

The idea that consciousness has something to do with the awareness of our mental states has a venerable tradition dating back to at least the writings of John Locke (Tye 1995, p. 5): "consciousness is the perception of what passes in a Man's own mind" and perhaps even to Aristotle (Güzeldere 1995, p. 335). This intuition has recently been given prominence under the name of the higher-order-thought theory of consciousness. Different versions of this theory differ as to the nature of the second-order state required. Armstrong (1980), like Locke, sees it as a perceptual state, a higher-order act of observing our first-order mental states. Rosenthal (1986) sees it as a more cognitive state, and Carruthers (1996) as a potential for being recursively embedded in higher-order states (see Güzeldere 1995). The basic insight behind these different approaches is that when one is conscious of some state of affairs (e.g., that the banana in one's hand is yellow) one is also aware of the mental state by which one beholds that state of affairs (i.e., one sees that the banana is yellow). There is something intuitively correct about this claim, because it is inconceivable that one could sincerely claim, "I am conscious of this banana being yellow" and at the same time deny having any knowledge of whether one sees the banana, or hears about it, or just knows of it, or whether it is oneself who sees it, and so on. That is, it is a necessary condition for consciousness of a fact X that I entertain a higher mental state (second-order thought) that represents the first-order mental state with the content X.

Of course, there is philosophical controversy about whether this characterisation can capture the whole phenomenon of consciousness or just an aspect of it.[11] We need only focus on the less controversial part of this theory, namely, that the higher-order mental state is merely necessary, although, in what follows we will occasionally explore the explanatory power of the stronger theory (that a higher-order thought is both necessary and sufficient for consciousness). To be safe we will pursue Carruther's "potentialist" version of the higher-order thought theory in more detail. Because it does not require actually entertaining a higher-order thought but only the potential for forming such a thought, it makes fewer demands on the cognitive complexity of routine conscious information processing than the other versions of this theory. This potentialist version is sufficient to explain why consciousness relates to explicitness, verbal expressibility, voluntary control, and so on.

Carruthers (1996) sees consciousness as the potential of our mental content that is recursively embedded in higher-order states. In other words, the content X of a knowledge state is conscious if it is recursively accessible to higher-order thoughts, such as knowing that I know that X. To form this second-order state one must explicitly represent the first-order knowing. For this, one in turn needs to represent the content explicitly, in particular its factuality, that is, "it is a *fact* that X." This is a necessary condition. It is not always necessary to have the first-order attitude and the self explicitly represented because these can be freely inferred from the factuality of the content as Gordon (1995) has pointed out in the context of simulation theory. Within one's own perspective – and that is all we are concerned with here – there is a one-to-one correspondence between what is a fact for me and what I know. Gordon speaks of ascent routines that allow us to go from descriptions of facts to knowledge attributions for oneself, for example, from "X is a fact" I can go to "I know that X." That means that once factuality is represented explicitly, explicit representation of attitude and self is also possible. Of course, other conditions may have to be met (e.g., the representation must be in a short-term memory store), but explicit representation of factuality (and thus all other aspects of content) is often all that is required.

In sum, on the weak version of the higher-order thought theory where potential access to higher-order thoughts is only a necessary condition, we can conclude that the explicit representation of self and attitude is necessary for conscious knowledge but sometimes only the explicit representation of factuality is all that is needed. On the stronger version, where access to higher-order thoughts is also a sufficient condition, explicit representation of self and attitude or factuality is sufficient for conscious knowledge. For us the critical implication of this view of consciousness is that the requisite higher-order states represent the attitude and the holder of the first-order state explicitly. This in turn requires explicit representation of the content of the first-order mental state. This means that to have conscious knowledge one must represent all three aspects of it explicitly (or be able to form such explicit representations). For example, to consciously know that the banana is yellow, I must explicitly represent that it is a present fact that the banana is yellow, that this fact is known, and I must be able to explicitly represent that it is I who know it. This analysis makes clear why most definitions of explicit knowledge involve consciousness: because it imposes the clearest, most extreme case of explicitness. It also puts us in a good position to understand why verbal access to knowledge (and other features to be discussed below) are tied to consciousness.

3.2. Verbalisation and directness of tests

In this subsection we wish to show why verbal access to knowledge is considered a sign of explicit, conscious knowl-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

edge, relating this to the important types of direct and indirect tests and the objective and subjective thresholds of perception.

Verbal communication (for transmitting information) proceeds by predication. A referring expression (or an ostensive gesture) is used to identify an individual (topic) and then further information about this individual follows. Hence, verbal report requires knowledge with explicit predication. An even stronger explicitness is necessary for the following reason. Linguistic information, unlike perceptual information, cannot be taken at its face value. As Gibson (1950) has emphasised, visual perception is highly reliable under most normal circumstances and thus can – barring the few visual illusions – be taken as veridical. Applied to linguistic information, this strategy would lead to a highly unstable knowledge base (Perner 1991, Ch. 4). For this reason, verbal information needs to be interpreted without being taken as true prima facie. Only after evaluation (checking compatibility with other available information) should it be accepted. A distinction must accordingly be made between "is a fact" and "not yet clear," that is, factuality has to be represented explicitly.

In research on implicit memory (Richardson-Klavehn & Bjork 1988) and subliminal perception (Reingold & Merikle 1988), a critical distinction is made between direct and indirect tests of knowledge. A *direct test* is one that *refers* to the fact in question. An *indirect test* does not refer to the fact in question, but the answer to some unrelated question or the response to some stimulus shows that some information about the fact must still be present. In both literatures, the fact in question is the spatiotemporal context of the presentation of a particular stimulus. The key methodological difference between implicit memory and subliminal perception is in how long after the presentation of the stimulus knowledge of this fact is tested (Kihlstrom et al. 1992). In implicit memory, the fact in question could be that a particular word was studied 10 minutes ago in the laboratory, and typically the word is consciously perceived at the time of study. Implicit memory is considered in more detail in section 5.2 below. In subliminal perception, the fact in question is whether a particular stimulus has *just* been presented. According to the normal approach (e.g., Holender 1986), perception is subliminal or implicit (Kihlstrom et al. 1992) if the participant performs at chance on a direct test of some aspect of this fact (because it was not consciously perceived), but the stimulus still affects processing indirectly.

Our analysis makes clear why performance on direct and indirect tests has something to do with implicit-explicitness and consciousness of the probed knowledge, provided the test questions are "bona fide," with participants saying that X is the case only if they have a representation stating that X is a fact. The analysis also makes it clear, however, that one cannot equate test performance with type of knowledge, because there is no guarantee that test answers are bona fide; participants might say that X is the case just on the basis of a feeling that that might be right.

Even knowledge without explicit predication can influence indirect test responses, because the test does not refer to the event in question. For example, after a brief (e.g., 10 msec) presentation of the word "doctor" or "table" followed (within, e.g., 50 msec) by a patterned mask (a technique for inducing subliminal perception), a clearly visible word (e.g., "nurse") or nonword (e.g., "nurge") is presented and observers must judge whether or not this item is a word; this lexical decision provides an indirect test of knowledge of the first word. Although the instructions refer only to the clearly visible word, it has been found (e.g., Marcel 1983a) that the identification of "nurse" is faster if the first word is semantically related (i.e., "doctor") than if it is unrelated ("table"). For this processing advantage to occur it is sufficient to take in only the property of the presented stimulus, that is, "doctor," without any representation that there was a particular event that had that property. The semantic processing triggered by the word form "doctor" will activate the semantic field of the medical profession, which then gives "nurse" a greater processing advantage than "table."

In contrast, a direct test refers to the event in question. There are different ways of making this reference. The question can refer to the event, for example "What was the word on the screen?" A bona fide answer "doctor" can be given only if the event has registered *as a fact.* So we see that bona fide performance on such a direct test requires explicit representation of factuality which, on Carruthers's higher-order theory of consciousness, is at least a necessary and possibly also a sufficient condition for consciousness. This provides a theoretical justification for using direct tests to assess conscious knowledge if all answers are bona fide. Unfortunately, there is no guarantee of this. Cooperative participants in our experiments try to give the best answer, and then even knowledge with implicit predication (far removed from meeting the criterion for consciousness) may help them give correct answers (based on guesses) on direct tests, a known problem in the field (e.g., Roediger & McDermott 1996).

Performance on indirect tests can be influenced by conscious knowledge as well as implicit knowledge lacking explicit predication. One could only infer the use of implicit knowledge without consciousness from the advantage in performance of an indirect over a direct test (even if non-bona fide answers are given on the direct test). This conclusion is warranted especially if performance on the direct test outstrips performance on the indirect test under conscious processing conditions so that any lingering issues about sensitivity differences (Shanks & St. John 1994) are eliminated (Reingold & Merikle 1993, p. 53).

Because direct tests do not typically involve reference to one's subjective mental state of seeing, Cheesman and Merikle (1984; see also Greenwald 1992) referred to the threshold conforming to this test as the "objective threshold": If the interstimulus interval between a stimulus (e.g., a word) and a mask is reduced so as to make perception more difficult, the objective threshold is defined by the interstimulus interval at which the participant performs at chance on a direct test of the nature of the stimulus presented. Our analysis, however, suggests that this might not reflect a single threshold because there are at least two significantly different ways of making such a reference (cf. Dagenbach et al. 1989). One is to stipulate that an event occurred, with the observer's task being to determine of which type the event was, for example: "What was the word on the screen?" This way of questioning puts the focus of the observer's mental search on finding a suitable property for an answer. A predication-implicit representation of the perceived property will serve that purpose.

A different way of phrasing the question is to stipulate a particular event type, for example, the occurrence of a

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

word; the observer's task is to decide whether or not it took place (i.e., to judge the existence or occurrence of a word). Marcel's (1983a, Experiment 1) query about whether a word (any word) was *present* or *absent* to determine the detection threshold appears to be of this kind. Here observers had to judge whether or not a word occurred. Such a judgement would require a predication-explicit representation of the perceived event. A mere representation of the property "word," without explicit predication of the observed event, would not provide a natural answer to the observer's mental search initiated by the presence-absence question. Several studies and replication attempts inspired by Marcel's work used the other approach for determining the detection threshold, for example, "Which colour word was it (one of four possible colours)?" (Cheesman & Merikle 1984) or "Was there a word or a blank?" (Dagenbach et al. 1989). In this case, a predication-implicit representation of the event type ("red" or "word" or "blank") provides an answer for the mental search. This may be one reason these studies had only partial success in replicating Marcel's original finding that detection (absence-presence) has a higher threshold (i.e., occurs at a longer stimulus onset asynchrony, SOA, between stimulus and mask) than graphic or semantic similarity judgements (also see Fowler et al. 1981).

There is also the possibility of formulating a direct test by referring to the target event as a perceptually experienced event: "What was the word *that you just saw?*" For a bona fide answer the stimulus event must be encoded explicitly as a *visually perceived event.* Without that encoding the observer can only answer "I didn't see anything."[12] Because reflection on one's state of seeing is required, this detection criterion corresponds to the "subjective threshold" introduced by Cheesman and Merikle (1984: 1986; see also Merikle 1992); that is, the point at which participants know they know what they saw.

This discussion was mainly intended to show that the known problems in this field can be formulated in our framework. The contamination of explicit (direct) tests by implicit knowledge and of implicit (indirect) tests by explicit knowledge has been debated particularly intensively in memory research. As a solution, Jacoby (1991) proposed his process dissociation procedure, which brings in voluntary conscious control as an arbiter. We will discuss the relation between the implicit-explicit distinction and consciousness and volition in the next two sections.

3.3. Procedural versus declarative knowledge and accessibility

The notion of procedural and declarative knowledge has been related to the implicit-explicit distinction by several authors. Karmiloff-Smith (1986; 1992) characterized as procedural the implicit knowledge that is severely limited in its accessibility to other parts of the system. Accessibility has been emphasised as the central factor in the distinction between procedural and declarative knowledge by Kirsh (1991). Squire (e.g., 1992) characterized the knowledge of the past that is typically impaired in amnesics as declarative memory (where declarative is considered largely a terminological variant of explicit memory or "knowing that"); he contrasted this with nondeclarative (implicit, knowing how) memory, which includes procedural memory (habits, skills, and conditioned reactions) but also memory of facts revealed by priming.

Our own suggestion is that at least four different dimensions are in play and need to be kept conceptually distinct: knowledge that is or is not contained in a procedure, declarative versus nondeclarative knowledge, accessibility, and implicitness versus explicitness. The goal, however, is to show that there are some necessary relations between these dimensions and the types of knowledge that form natural clusters: procedural knowledge tends to be implicit and hence inaccessible, whereas declarative knowledge involves quite explicit representation of its content, and hence tends to be conscious and accessible for different uses.

To some, implicit knowledge may simply mean inaccessibility. Apart from being an arbitrary conceptual stipulation, this definition of implicitness also lacks precision. Inaccessible in what way? All knowledge must be accessible in some way or it would not qualify as knowledge (on views like those of Millikan 1984; Dretske 1988); in any case, there would be no evidence that there was any knowledge at all. Our framework indicates how the implicitness of different aspects of knowledge makes it inaccessible in different ways, as indicated in our discussion in section 3.2 on direct and indirect tests and verbalisability, and in our treatment of procedural knowledge, which we now discuss.

The procedural-declarative knowledge distinction was introduced in artificial intelligence (McCarthy & Hayes 1969; Winograd 1975) and later taken over in psychological modelling by Anderson (e.g., 1976). It concerned how best to implement knowledge: Should one represent the knowledge that all men are mortal as a general declaration "for every individual it is true that if that individual is human it is also mortal"? Whenever knowledge of a human individual was introduced in the database this general information would be consulted to infer by general inference rules that that individual must also be mortal. The alternative would be to have a specialised inference procedure: "Whenever an individual is introduced that is human, represent that that individual is mortal."[13]

Now we can see in what sense declarative knowledge is explicit. It represents explicitly that the regularity "if human then mortal" is predicated of individuals and its general application to every individual is also marked. Moreover (if the database provides the requisite expressive power), it states that this regularity is a fact. In contrast, the procedure that adds "is mortal" to every human individual it encounters also knows something about this regularity but its knowledge is implicit in its application; its generality is implicit in the fact that it is applied to every individual encountered. No distinction, however, is made in the system that represents that it is applied to individuals and to every individual. The analysis also brings out the intuitive meaning of declarative knowledge as knowledge that declares what is the case (e.g., Squire 1992, p. 204: memory whose content can be declared) because it represents explicitly that something is a fact. Nondeclarative memory can be given precision in our analysis either as the stronger form of knowledge that does not make predication explicit or as a weaker form of knowledge that makes predication explicit but leaves factuality implicit.

The implicit nature of procedural knowledge also makes it clear why it has limited accessibility. For example, the implicitness of the procedural representation of the fact that all humans are mortal does not allow the distinction between whether this rule applies to a current case and my

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

thinking about the rule. To separate these two cases one needs some internal distinction that (explicitly) represents whether or not the rule applies. Then one can distinguish whether one is just thinking about the rule without it actually applying, or whether one is thinking about it because it applies. Moreover, there is no way of checking the adequacy of procedural knowledge. Such a check requires explicit representation of factuality to represent the result of the inference as a hypothetical possibility, for comparing it with other available evidence.[14] All this puts a severe limitation on the usability of procedural knowledge.

The advantage of procedural knowledge is its efficiency. Procedures need not search a large database because the knowledge is contained in the procedures. Knowledge that resides in the application of a procedure leaves predication and factuality implicit. As a result, it is limited in its accessibility in a way that has been claimed for modularity (Fodor 1983); modular knowledge, for example, applies only to a specific input modality; it cannot use knowledge from other domains. Implicitness of procedural knowledge is accordingly a natural basis for modularity in our input modalities, which do not require fact explicit representation. In this context modular knowledge can be called implicit. However, implicitness is a less natural ally of modularity in the case of central processes (Fodor 1987b, "modularity gone mad").

Modular or quasi-modular central conceptual processes have been proposed by Cosmides (1989) for reasoning processes that use a cheating detector module. Sperber (1996) considers quasi-modularity a general feature of central cognition. Smith and Tsimpli (1995, Ch. 5) posited a quasi-modular central language module to explain isolated highly developed foreign language ability in an otherwise handicapped individual. The central language module is not the same as the usual linguistic input processing module because it is not used to converse in different languages but to translate playfully from one language into another. Such central modules are unlikely to operate purely procedurally without explicit predication or factuality. This is very clear in Leslie's (1987; 1994) theory-of-mind module, proposed to explain the relative ease and speed with which children develop a theory of mind. A theory of mind does not just process factual information. It must represent the content of people's beliefs and desires, hence explicit representation of factuality is required. Modular knowledge in this sense clearly cannot be implicit in the sense defined in this paper.[15]

In sum, knowledge contained in the application of a procedure (procedural knowledge) is active and efficient, but it leaves predication and factuality implicit; hence it is non-declarative and limited in its range of applicability (hypothetical reasoning, checking validity) and far from being accessible to consciousness. In contrast, knowledge that states its predication and factuality explicitly cannot be contained in the use of a procedure. It thus loses efficiency but becomes more flexible, to be used in hypothetical reasoning, evaluation of truth, and conscious awareness. The distinction between procedural and declarative knowledge is a good basis for understanding why voluntary control of action is tied to explicitness and consciousness.

3.4. Voluntary control

The dominant philosophical view of what differentiates our intended actions, for which we are responsible, from other movements is that those actions must be caused by our desires and beliefs (Davidson 1963). Heyes and Dickinson (1993), discussing whether animals act or just respond, argued that intentional action – unlike responses – must be based on an understanding of why one performs them, that is, one must represent the goal one pursues and the fact that the action leads to that goal. Searle (1983) even argued that intentional action must be causally self-referential: one must intend that the action be caused by one's intention.

A useful model for this phenomenal distinction between automatic (responses) and controlled, or willed action is that of Norman and Shallice (1986). It distinguishes two levels of control. *Horizontal strands* operate at the level of implementing schemas, which consist of complex conditional action tendencies (productions like in Anderson's 1976 ACT model) with automatic control through activation by triggering stimuli and mutual inhibition of simultaneously triggered schemas (*contention scheduling*). [See also Anderson: "Is Human Cognition Adaptive?" *BBS* 14(3) 1991.] *Vertical strands* of control come from the *supervisory attentional system* (SAS, a close relative of the central executive; Baddeley 1986). The two control systems are supposed to capture the phenomenal distinction between automatic responses and intentional action as well as explaining why a particular set of actions becomes difficult for patients with problems of voluntary control (e.g., patients with frontal lobe insult). These "SAS tasks" are typically (1) the setting up of new action schemas upon task instructions, (2) monitoring of novel or dangerous actions, or (3) the inhibition or monitoring of interfering action schemas.

Action schemas or productions are complex versions of responses to stimuli. They incorporate procedural knowledge about event contingencies in the world that (as discussed in sect. 3.3) leave predication of these regularities and factuality implicit in their application. The stimuli that trigger them can be declarative or nondeclarative representations of features of the environment or internal states. The control exerted at the level of contention scheduling, as well as that exerted by the SAS, is in terms of boosting or inhibiting the activation of schemas. For example, in order to ensure that a single schema produces coherent action the dominant schema might get its activation boosted even further at the cost of the activation of less dominant schemas.

We suggest that contention scheduling directs this control purely on the basis of the schemas as representational vehicles (the amount of activation is a feature of the schema as a vehicle, not of its representational content). In contrast, the SAS directs its control on the basis of the schemas' representational content. In support of this contention one can show that such content-oriented control is necessary for the "SAS tasks" listed by Norman and Shallice. In a version of the Wisconsin Card Sorting test for children, a three-year-old child (like a frontal lobe patient) who has learned to sort cards by colour must now sort the same cards according to a new rule (e.g., the shape of symbols on the card). Without SAS, the once-learned colour sorting rule is dominant and will suppress execution of the new rule. Three-year-old children, even though they know the new rule and can verbally state it, will perseverate, sorting according to the old rule (Zelazo et al. 1995), as frontal lobe patients tend to do on the traditional test (Shallice 1988). For the SAS to be of use here, it must boost the new schema and inhibit the old, dominant one. This cannot be done on the basis of vehicle

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

features such as amount of existing activation or strength (too many weak schemas would be boosted), the SAS must be able to address the new schema by its content, that stimulus-response sequence that the new rule requires (see Perner 1998, for discussion of other SAS tasks).

Controlling schemas via their content requires the representation of that content. To avoid confusion, this content must be explicitly marked as being not factual (i.e., explicit representation of factuality), but something that is desired or intended (explicit representation of attitude). This means that the SAS must be (or contain) a second-order mental state (one that represents desires), which is an important prerequisite (or even a sufficient condition) for being a conscious state according to the higher-order thought theory of consciousness. This analysis hence suggests that the need to represent content and attitude explicitly distinguishes controlled or willed action from automatic action. We can identify intentional action with action (be it automatic or willed) that is in line with the explicit representations of the SAS (under control). If automatic action contravenes those representations then it is experienced as an unintentional lapse or "slip of action" (Reason & Mycielska 1982).

This analysis also makes it clear why willed action is conscious – because it is based on a second-order mental state. With this we have a theoretical justification of why voluntary control is used as a criterion for consciousness in the quite different areas of research on implicit memory and subliminal perception. Note, however, that not all aspects of the content of a schema need be explicitly represented to allow control by the SAS – only enough aspects to indicate that the action of the schema is desired. Only those aspects of the content that are explicitly represented will be conscious; the rest may in principle embody knowledge that the person is not aware of having, and whose details of application they could not control. Our argument requires a conscious representation to be made by the SAS (e.g., "I want [it to be the case that] I play 'Für Elise' on the piano"), but the overlap in content between this representation and a body of knowledge (namely, about piano playing) could allow that knowledge to apply, even if the factuality of the knowledge is not explicitly represented; that is, a fully explicit representation in the SAS can coexist with implicit representations in a knowledge base. We will see an example of this in section 4.4.

Jacoby's (1991) process dissociation procedure uses voluntary control of knowledge to provide better estimates of implicit (unconscious) or explicit (conscious) memory. The procedure can be used not only for memory but also for subliminally presented information (Debner & Jacoby 1994). One critical part of this procedure is the exclusion condition, in which participants in an indirect test of memory (e.g., to complete word stems) are instructed not to use words that were presented in a list. Unconscious knowledge, in particular, knowledge that leaves predication implicit (e.g., the word form "butter" of the word that was on the learning list), can influence the indirect test and escapes exclusion in the exclusion test, because the word form does not fall under the description "word on that list." So, the number of words from this list that are used as an answer, despite instructions, is a better indicator of implicit memory than performance on the indirect test without exclusion instruction, because on the indirect test there is no control for participants using words they can remember explicitly.[16]

3.5. Summary

Our analysis of the aspects of knowledge that are represented explicitly and those that are left implicit provides a basis for relating different criteria that have been brought to bear on the implicit-explicit distinction. Knowledge that represents its content, its attitude, and its holder (self) explicitly is on the higher-order thought theory conscious. Explicit representation of factuality might be sufficient: because from being a fact, knowledge can be inferred. Explicit representation of predication (and often of factuality) is required to refer in verbal communication and thus a link emerges between direct tests (where reference is made to the known fact) and explicitness and consciousness. Similarly, procedural knowledge leaves predication implicit in its application. It accordingly remains unconscious. Declarative knowledge represents predication and factuality explicitly, thus qualifying for conscious access. Automatic action is based on schemas (productions) that, like procedural knowledge, leave predication implicit, while controlled action (SAS) represents the content of these schemas explicitly, together with the attitude. Willed action is therefore conscious, whereas automatic action can remain unconscious. This justifies the use of voluntary control to help distinguish conscious from unconscious elements in task performance.

4. Outline of potential application to research areas

4.1. Visual perception

Visual information is not processed in a unitary way. At least two functionally different systems exist. Traditionally it was thought that the functions were for the perception of objects and the perception of the spatial relations between these objects ("what" versus "where;" Ungerleider & Mishkin 1982). Recently, Milner and Goodale (1995) have moved from a distinction in terms of encoding different aspects of the visual array to either forming a perceptual representation ("what" there is) versus exerting visuo-motor control ("how" to act). This reconceptualisation has been prompted in large part by functional dissociations in brain-injured patients and normal people (e.g., Milner & Goodale 1995; Rossetti 1998). As an example, we describe a series of experiments by Bruce Bridgeman on the induced Roelofs effect.

Bridgeman (1991; Bridgeman et al. 1997) reports that for human observers a stationary dot within a rectangular frame appears to move opposite to a movement of the frame. After a brief exposure to this apparent movement, the display vanished and the observer had either to indicate verbally at which of five marked locations the dot had been after the movement or to point to the location of the dot. In their verbal responses all observers were susceptible to the illusion and reported the dot's last location as having moved opposite the frame's movement. In contrast, only half the observers were susceptible to the illusion in their pointings; the other half pointed quite accurately to the dot's actual location. Bridgeman interprets the results as showing the dissociation between a cognitive (perceptual) system used for verbal report and a system for visuo-motor control that steers the pointing finger.

This interpretation can be refined within our conceptual framework. Visually guided behaviour can be procedural

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

and nondeclarative; it does not need a distinction explicitly between facts and nonfacts. It is a system that registers object features in egocentric space and everything that is represented is a fact. An interesting question is whether predication needs to be represented explicitly. It seems that the object one grasps does not need to be represented as a reidentifiable individual. Representation of its visible features suffices[17] as Campbell's (1993) analysis shows that orienting oneself in relation to landmarks can be done in a pure feature placing system without the necessity of conceptualising the landmarks as physical objects that have those features. So, no predication of the visible features to the objects that have them needs to be represented. This still leaves the question, however, of whether the visible object features need to be predicated of the spatial positions, that is, "dot-ness in position x, y, z," which amounts to predicating the feature "dot-ness" of that position. Or is it sufficient simply to have a conjunction of feature and position? A plausible answer might be that a mere conjunction is sufficient if only a single object needs to be tracked. Then the predication of feature to position can remain implicit in the tracking. For keeping the position of a second feature in mind while tracking the first, explicit predication is required. We know of no data that speak to this issue,[18] but the question of whether visually guided action leaves only factuality and time or also predication implicit is testable.

In contrast to visually guided behaviour, to give a verbal response is to make a judgement that that is where the dot really is. The information in this system needs to represent predication and factuality explicitly. As these are preconditions for consciousness, this explains why the information used for the verbal response is what is consciously experienced. The analysis also makes clear a certain ambiguity in the pointing condition. Pointing is a movement of the finger to the target (a visually guided movement); but it is also a declarative act that states what is the case. The bimodal distribution could be due to this ambiguity. From our analysis it follows that if the instructions are not to point but to move one's finger to touch the dot, then no observer should be susceptible to the Roelofs effect. Bridgeman (personal communication) carried out this condition and obtained the predicted results.

Bridgeman's experiment illustrates the other interesting property of the visuo-motor system: that its information persists for only a few seconds. When the response is delayed for eight seconds, all observers show the Roelofs effect just as in their verbal response (and this also holds for the condition where observers had to move their finger to the target; Bridgeman, personal communication). Representations that do not mark factuality and time are only useful to represent the here and now, as they do not differentiate what is a fact (here and now), what is not a fact but merely a hypothetical assumption, or what was a fact but is no longer (see Perner, 1991, for developmental convergence of the ability to represent hypothetical scenarios and to represent change over time). So, because the visuo-motor system leaves time and factuality implicit, it can only update its information about the current state of the environment but cannot keep track of past states of affairs and compare them with the present one. For this, factuality and time must be represented explicitly (see alsoWong & Mack 1981).

In sum, what these results demonstrate is that there are two visual information processing systems. One is identified neurophysiologically with the dorsal path from the primary visual cortex (V1) to the posterior parietal cortex (Milner & Goodale 1995). Its information is unconscious, it cannot be used for statements (verbal or gestural) about the world, it is not susceptible to certain illusions, and it is used for action in the world but is of limited duration. Our interpretation is that this system leaves factuality and time implicit (and perhaps also predication – see above). The other system is identified with the ventral path from V1 to the inferotemporal cortex. Its information is conscious, susceptible to illusions, and used for statements about the perceived world and for action in the world after some delay. This system represents predication and factuality explicitly and thus makes its content accessible to consciousness (see also Aglioti et al. 1995; Gentilucci et al., in press; Milner & Goodale 1995, Ch. 6; Rossetti 1998).

The spared capacities in blindsight and blind-touch patients (tactile analogue to blindsight; Paillard et al. 1983) depend on similar parametric variations. Marcel (1993) reported that blindsight patient G.Y. was able to detect an illumination change in the blind field better when the response was made quickly than when it was delayed by 2 or 8 seconds, when the response consisted of an eye blink (interpretable as a nondeclarative response) than a verbal "yes-no" (a declarative comment), and when the patient was invited to guess than when instructed to give a firm judgement (where bona fide responses require judgement explicit representation). Marcel also found that people of normal vision responded to near-threshold changes in illumination in the same way as blindsight patients. That is, in people with normal vision, detection was better when responses consisted of an eye blink rather than a "yes-no" verbal response, and when people were invited to guess rather than make a firm judgement.

A particularly interesting point about the last result is that the response shift from judgement to the guessing condition consisted not of a criterion shift to saying "seen" more often, but of an increase in discrimination accuracy (an increase in hit rate *and* decrease in false alarm rate). A shift in criterion towards "seen" responses would be expected if the stimulus was encoded *explicitly* as a fact about which one is uncertain in one's judgement. Then being given leave to guess would simply lower the rejection criterion resulting in an increase in the willingness to say "yes." In contrast, when a stimulus is encoded as a fact only implicitly, there is a representation "illumination change" but no information as to whether or not it occurred, or whether it occurred on the current or an earlier trial. Thus, there is no proper information for a judgement (hence low detection accuracy). With leave to guess, however, one is free to let oneself be influenced by the fact-implicit information that happens to be correct, which results in higher detection accuracy.

4.2. Memory

Memory has many different facets. To help focus our discussion, we distinguish the wider use of memory as the availability of information acquired in the past (e.g., remembering/still knowing that $2 \times 2 = 4$) from the narrower meaning of memory as the availability of information about events in the past acquired in the past. As a concrete example, we use the typical memory experiment in which one is read a list of words, among them the word "butter," and we look at the consequences when various aspects of this

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

event are represented explicitly or left implicit. The consequences we consider are in terms of memorial state of awareness, retrieval volition, and test responses.

As the first possibility, we consider strong implicitness. At learning, the word "butter," designed to represent the fact that "the word 'butter' occurred on the list," is stored so that only the word form "butter" is represented explicitly and all the rest is left implicit. This supports no particular memorial state of awareness. It could support a "feeling of familiarity" if that word had been encountered the first time on that list. This representation cannot be accessed voluntarily, and is not used bona fide in any direct test because no reference to any particular occurrence can be made. It can influence indirect tests, however. The mere presence of the word form "butter" can enhance the likelihood of answering a request to list dairy products with "butter." It could also account for participants reporting "butter" on an exclusion test without any accompanying feeling of familiarity (Richardson-Klavehn et al. 1994).

It is also likely that there are cases where it is not just the word form "butter" that has been represented, but also the perceptual details by which that word form was perceived. That is, a representation of the conjunction of various contextual features is formed, but this feature complex need not be predicated as having occurred on the list. Such a representation could enhance perceptual identification and produce familiarity effects without supporting recollection (e.g., Jacoby & Dallas 1981). Such a representation could also be involved in the "mere exposure effect," in which exposure to a stimulus, for example a novel shape, can lead to high affect ratings for the stimulus in the absence of recollection of having seen it before (Bornstein 1989; Gewei & van-Raaij 1997; Zajonc 1968).

When the occurrence of the word "butter" is explicitly predicated, "the word 'butter' occurring on that list," then it can come under direct voluntary control because now reference to the particular event of being on the list is possible. As a consequence, performance on a direct test can be better than on an indirect test (Reingold & Merikle's 1993 control for differences in test sensitivity). However, voluntary control remains an educated guess and does not result from a considered judgement, because the occurrence is not represented as a fact.

Explicit representation of the occurrence as a fact makes the event accessible under the description of being a fact and participants can now give a considered judgement that the word "butter" is part of that list. With explicit representation of time, participants can then also judge that "butter" occurred at a particular reading of the list in the past. They can experience memory of a past event. It can be a conscious experience of memory of the past according to the higher-order thought theory, because explicit representation of factuality entails a higher-order thought about one's knowledge. However, even with such a representation participants may remember no details of seeing/hearing the item.

An important next step comes with explicit representation of the experiential source of one's knowledge: "I know that 'butter' was on the list because I saw it there." Only such encoding – encoding of having been in direct contact with the known event – constitutes *genuine episodic memory* according to Tulving (1985; Perner 1991).[19] Tulving (1985; and later others, such as Gardiner 1988) distinguished two types of recognition responses: those accompanied by simply an experience of knowing that the item occurred earlier in the context of the experiment ("K" responses), and those based on truly remembering the prior experience of the item ("R" responses).

"K" responses may arise for various reasons: because the word form "butter" is encoded predication implicitly and simply comes to mind readily (whether the participant does give a positive recognition response depends on his theory of why the word came to mind) or because a predication explicit representation has been formed and hence the participant guesses that the word had been on the list. In both cases, the participant may give a "K" response with low confidence. On the other hand, if the participant experiences strong familiarity when he comes across the word "butter," he may give a "K" response with strong confidence. However, in all these cases there is no genuine knowing that "butter" was on the list, just guesses that carry more or less conviction. Researchers in the field (Conway et al. 1997) have now started to give participants a choice between "K" responses and "guesses." This may separate predication and fact-implicit knowledge from knowledge that represents the factuality (and past-ness) of the event in question explicitly. Unlike "guesses," "K" responses should not just be produced but produced as the reflection of a fact. "R" responses differ from "K" responses in that they need be seen not only as reflecting facts but also as products of one's direct experience.

Table 2 summarises the different levels of explicitness, and the memorial state of awareness, voluntary control, and kind of test performance they support. Our analysis yields distinctions that map reassuringly onto distinctions that have emerged from the empirical literature. In particular, it can address the distinction between *retrieval volition* and *memorial state of awareness* (Richardson-Klavehn et al.

Table 2. *Relation between type of representation and type of memory*

Laid down representation of fact that Fb	Memorial state of awareness	Retrieval volition	Reference by:	Recognition test response
Property "F"	none	involuntary	nothing	correct guess.
Compound "F-X"	feel of famil.	– " –	nothing	recogn. by famil.
Predication "Fb"	– " – .	direct vol.	"part of list"	– " –
Factuality				
+ Time "Fb happened"	knowing past	– " –	"was on list"	"K" (past event)
Origin "I experienced Fb"	remembering	– " –	"remember!"	"R"

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

1996; Schacter et al. 1989); it honours the distinction between "implicit" memory and the distinction between "know" and "remember" judgements as two kinds of explicit memory in the spirit of Tulving's (1985) original distinction, where "know" judgements are supposed to cover "knowledge of the past" and "remember" judgements memories of experienced events as experienced (Perner 1990). This analysis indicates that both "R" and "K" count as declarative knowledge (both involve explicit predication) and familiarity can be purely procedural (predication left implicit).

4.3. Development

In our framework there is no simple dichotomy between implicit and explicit knowledge. This owes much to Karmiloff-Smith's (1986; 1992) insistence that the basic dichotomy should be embellished by further levels of explicitness. It is reassuring that our framework unfolds logically from the conceptual analysis of knowledge and yields a plausible correspondence to Karmiloff-Smith's empirically motivated classification. Her initial level (I) of implicit knowledge, where the information is only *in* the system, maps onto procedural knowledge, which leaves predication implicit. Her first level (E1) of explicit knowledge results from a redescription of the original information encoded in procedural format, so that the information becomes information *to* the system, useable by different parts of the system. This maps onto knowledge that makes predication explicit (and can thus be referenced flexibly by different user systems) but leaves factuality implicit. At the next level of explicitness (E2), the knowledge becomes conscious and at the final level (E3) it is also verbally expressible. The once clear progression from E2 to E3 has later been collapsed into a level E2/3 (1992, p. 23), owing to the lack of a clear empirical demonstration of such a progression. The level E2/3 corresponds to knowledge that makes factuality (and source) explicit. Moreover, because explicit factuality tends to make knowledge conscious and verbally accessible, our analysis actually suggests the merging of the original levels 2 and 3.

Whereas Karmiloff-Smith's research emphasises how implicit knowledge becomes increasingly explicit with development, dissociations between two competing knowledge bases have also been found – dissociations reminiscent of those in visual perception (e.g., Clements & Perner 1994; Diamond & Goldman-Rakic 1989; Goldin-Meadow 1993). Goldin-Meadow et al. review studies that show that the acquisition of concepts of quantity (Piaget & Inhelder 1974/ 41) can be more advanced in children's gestural comments than in their verbal responses. One of the interpretations of this finding was (Church & Goldin-Meadow 1986) that the multidimensional spatial medium of hand gesture makes it easier to express novel ideas than the unidimensional temporal medium of linguistic expression. However, one can think of the gestures as spontaneous (mostly unconscious) concomitants of the thinking process. In that case the earlier emergence of advanced knowledge might be the sign of thoughts about reality that have not yet been recognised as being about reality (implicit factuality). This interpretation fits a parallel finding in children's developing "theory of mind."

Clements and Perner (1994) reported that the understanding of false belief emerges in children's visual orienting responses as early as 2 years and 11 months, a year earlier than in their verbal responses to questions. Children are told enacted stories in which the protagonist does not see how his desired object is unexpectedly transferred from one location(A) to another (B). Children in the interesting period around 3 years of age answer the question about where the protagonist will go to get his object wrongly by pointing to the current location of the object. However, a majority of these children look (visual orienting responses) in anticipation of the protagonist at the empty location where the protagonist mistakenly thinks the object is.

Further research (Clements & Perner 1997) indicates a remarkable similarity to the observed dissociations between the two visual systems (see sect. 4.1). When instructed to move a welcoming mat for the mistaken story protagonist who was on his way to get his object, children who move the mat spontaneously tend to move it correctly to where they think the object is (A), whereas children who need prompting (thus with some delay) move it to where the object actually is (B). There seems to be a stage in children's developing understanding of belief where two different knowledge bases dissociate. One of them is a more accurate and developmentally advanced knowledge base (by analogy with the dorsal visual path) that supports only nondeclarative action (looking and moving a mat) carried out without delay (spontaneous mat move), while a less accurate and less developmentally advanced knowledge base (analogous to the ventral visual path) is used for declarative responses (verbal and pointing) and delayed action (prompted mat moving). We do not know, of course, whether the more advanced knowledge is conscious and the other unconscious, since one cannot ask 3-year-old children to report on such a distinction, but otherwise the similarities are remarkable.

Such a similarity between dissociations in processing visual information about the environment and understanding another person's false belief suggests that the characteristics of the two types of knowledge are not determined primarily by the brain regions in which the information is processed (dorsal vs. ventral path) but by more general functional differences that apply to visual information processing as well as a theory of mind. Our analysis shows how these functional distinctions could arise from these aspects of knowledge that are represented explicitly. An interesting speculation about functional differences in the theory-of-mind case is that the explicit understanding comes with (something of) a real theory, that is, a causal understanding of belief formation and how belief determines action. In contrast, the implicit understanding of where the protagonist will go may be based on abstraction of situational regularities. Within our framework this assumption gives a quite coherent picture of the existing data and leads to new, testable predictions (Perner & Clements, in press).

One can learn that certain events tend to go together and form a typical sequence. Such filtering of statistical patterns of possible combinations does not need representation of individual events and inferences from individual events to all possible events. Rather, it is a process of pattern formation and recognition for which connectionist systems are good (e.g., to classify different feature patterns into letters; e.g., Bechtel & Abrahamsen 1991). The combinations of letters encountered in artificial grammar tasks have a similar effect and can be particularly well modelled by connectionist networks (Dienes 1992). Although individual in-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

stances shape the connections between units and hence the association between the properties these units represent, there is no representation of the individual instances.[20] Connectionist work also shows that such pattern generalisation leads to pattern completion. If many elements of a typical pattern are present, the network tends to generate representations of the missing bits. This is important, because such pattern completion processes can produce expectations of what is to come on the basis of what has happened so far. For us, the important implication is that such associative expectation is possible without explicit predication. [See also Pessoa et al. "Finding Out About Filling-In" *BBS* 21(6) 1998.]

This makes it possible to anticipate correctly where the protagonist will go to get the desired object in our false belief stories without explicit predication to a particular occasion, namely, without representing that he will go there. According to our discussion, such a representation of the mere event form "protagonist going to location A" and hence, "protagonist at location A" as part of a pattern completion process can guide visual orienting responses and spontaneous actions because it can trigger an existing action schema waiting to be executed. It cannot be used for communication because it fails to be predicated of an individual event that can be reidentified across mental spaces explicitly marked as "facts," "anticipation," or "verbal description." It cannot sustain uncertainty, as it does not support a self-reassuring check about where the protagonist will come down because without explicit predication there is no representation stating that he will go anywhere. This is the pattern of results we observed in the precociously correct responses: they were high only in spontaneous action and visual orienting responses.

In contrast, a theory of belief goes beyond mere generalisations of observed regularities and constitutes genuine causal understanding of the underlying processes (see Gopnik 1993 and Perner 1991 for indications of theory use). Causal understanding cannot be achieved by mere pattern matching and pattern completion but must use explicit predication because causal reasoning supports counterfactuals (Lewis 1986; Salmon 1984). Counterfactual support means one understands that if the conditions had been different, the result would have been different; such reasoning requires different mental spaces for contrasting the actual facts with their counterfactual oppositions. For these reasons, responses based on a causal theory of belief should also be accessible to communication (answers to questions) and be robust against doubt (hesitating action).

One can accordingly predict that implicit knowledge should be shown primarily in the situation described above, where the correct response can be based on situational, behavioural regularities, such as "people look for objects where they last put them, where they last saw them, where they told someone to put them," and so on. In the traditional scenario all these regularities – if they apply – point to the same, correct answer "A." In a variant scenario (Perner et al. 1987) the protagonist, who has put the object into B, tells a friend to move the object from B to A, but the friend forgets. Here, behavioural regularities give different predictions. "Last seen" or "where put" indicate location B while "told to put" indicates A correctly. Hence signs of implicit understanding should be reduced in this scenario. Indeed, Clements (1995, Ch. 5) reports that children show fewer orienting responses to location A than in the traditional scenario. In contrast, their verbal responses show little difference in the two scenarios, replicating the original result by Perner et al. (1987). This is to be expected if explicit responding is based on a causal understanding of belief formation.

Another prediction is that verbal explanations of why the protagonist believes the object is still in location A (in the original scenario) in contrast to observing behavioural regularities (seeing the protagonist look for the object in A) should affect implicit and explicit understanding differently. Causal explanations should primarily affect explicit understanding, whereas observing regularities should have a stronger effect on implicit understanding. The role of explicit understanding in this prediction has been tested. Clements et al. (1997) report that causal explanations affect verbal responses but the observation of regularities does not. The corresponding data on visual orienting responses or action responses are not yet available.

4.4. Artificial grammar learning

Our framework also elucidates the different ways in which knowledge can be implicit in the standard implicit learning paradigms. The paradigm explored most thoroughly in the implicit learning literature is artificial grammar learning (see Berry 1997 and Reber 1989 for overviews). In a typical study, participants first memorize grammatical strings of letters generated by a finite-state grammar. Then they are informed of the existence of the complex set of rules that constrains letter order (but not what they are), and are asked to classify grammatical and nongrammatical strings. In an initial study, Reber (1967) found that the more strings participants had attempted to memorize, the easier it was to memorize novel grammatical strings, indicating that they had learned to use the structure of the grammar. Participants could also classify novel strings significantly above chance (69%, where chance was 50%). This basic finding has now been replicated many times. So participants clearly acquire some knowledge of the grammar under these incidental learning conditions, but is this knowledge implicit? We will now analyse the case of artificial grammar learning theoretically and empirically in terms of the different aspects of being a fact or being knowledge that can be made explicit, or left implicit, according to our previous analyses. (See also Dienes & Perner 1996, who explore whether participants represent the property structure of a grammar implicitly or explictly, an issue not dealt with in the following.)

4.4.1. Predication. When participants learn the structure of an artificial grammar by exposure to the exemplars, they may not explicitly represent the particular grammar to which the properties are predicated. Consider a person who uses the mental rule that "M can be followed by T." This statement represents the fact that, according to the grammar one was trained on 10 minutes ago, M can be followed by a T. Yet, the fact that it is a particular grammar which has this property is not explicitly represented because there is nothing in the expression "M can be followed by T" whose function it is to covary with that fact. This fact can be made explicit by forming the mental expression: "*g* has the property that M can be followed by a T," where *g* denotes a particular grammar (e.g., the grammar that I was just being trained on). The critical feature here is that different properties, such as "my having just been trained on"

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

and "being a grammar in which M can be followed by T" can both be predicated of *g*. This extended expression makes the implicit predication of "M is followed by a T" of a particular grammar explicit, because the whole expression does have the function of covarying with the fact that the identified particular grammar is characterized by the property in question.

Whether participants represent the individual grammars and the predication relationship explicitly can be revealed by the *volitional control* that participants have over the application of their knowledge. Consider a test of volitional control given to participants by Dienes et al. (1995). Participants were given 7 minutes to memorize exemplars generated by one grammar, and then another 7 minutes to memorize exemplars involving the same 6 letters generated by a second grammar. Participants were then informed that two grammars were involved and given a test in which a third of the items followed the first grammar (but not the second, e.g., xmxrtvtm), a third followed the second grammar (but not the first, e.g., xmvrxrm), and a third violated both grammars (e.g., xmtvvxrm). Participants were asked to choose items that followed only one of the grammars; half the participants were asked to endorse only the items consistent with the first grammar; the other half only the items consistent with the second grammar. Participants were perfectly able to distinguish the grammars at the usual performance level in such tasks and showed no tendency to endorse the grammar they were asked to ignore. How could this performance be achieved?

One way to succeed in such a test is to have direct volitional control over one's knowledge, in the sense that one can decide to use or not to use *it* because *it* has been explicitly labelled as the particular body of knowledge one wishes to use or not use. That is, we assume that for direct control it is necessary to represent the individual grammar explicitly. There are other ways of controlling which body of knowledge to use, however, that do not require such explicitness. For example, Whittlesea and Dorken (1993) argued that participants could distinguish different grammars by familiarity. One account of the Dienes et al. (1995) results along these lines is that the choice of grammar can be made by means of a compound property (*in-context-A,-M-can-follow-T*). Context A could be, for example, a particular time at which a string was studied . If context A is reinstated by task demands or imagination, the knowledge of a particular grammar can be isolated (through association) without having to predicate these properties explicitly to any particular grammar.

Even though this scenario of indirect control over particular grammars without explicit representation of the grammar is often possible or even plausible, there may be situations in which one can plausibly decide that volitional control was actually mediated (at least in part) by explicitly representing the individual grammar. For example, if, with a sufficiently sensitive test, measures of familiarity (such as ratings, speed of stimulus identification) do not predict classification response, then these alternative scenarios (that do not represent the individual explicitly) are not supported. Buchner (1994) in fact found that grammaticality judgements were not related to speed of identification. If this type of observation is supported, it follows from the volitional control experiments that participants do represent the individual grammar (and the predication relationship) explicitly. Of course, as we noted earlier (sect. 2.1.3), the presence of knowledge in which the predication relationship is represented explicitly does not rule out the possibility that there is further knowledge on the same topic which is predication implicit.

4.4.2. Reflection on attitude. To clarify how explicitly participants can reflect on their knowledge, it is necessary to be clear about *what* piece of knowledge participants may be reflecting on (e.g., Shanks & St. John's 1994, information criterion). We distinguish two different domains of knowledge. The first we call *grammar rules.* These are the general rules of the grammar that the participant has induced, for example, "M can be followed by T." The second domain pertains to the *ability to make grammaticality judgements.* This arises when the grammar rules are being applied to a particular string and it pertains to the knowledge of whether one can judge the grammaticality of the given test string independently of knowing that one knows the rules one brings to bear for making this judgement.

Knowledge of artificial grammars and of natural language may differ. We seem to lack explicit knowledge of the grammar rules both of English (we cannot represent *any* sort of attitude towards most rules of English grammar, so such rules are at least attitude-implicit) and of the quickly acquired artificial grammars. In contrast, we are fully aware and have explicit knowledge of our ability to judge the grammaticality of English sentences. We lack this sort of explicit knowledge of our ability to judge the nonsense strings produced by an artificial grammar. (We may lack it in the early stages of learning a first or second language as well.)

Various relationships between the knowledge of rules and grammaticality judgements are possible. Reber (1989) showed that people do not use the rules to respond deterministically; that is, when retested with the same string, participants often respond with a different answer. Extending this argument, Dienes et al. (1997) argued the data best support the claim that participants match the probability of endorsing a string as grammatical to the extent to which the input string satisfies the learned grammatical constraints, and that this probability varies continuously between different strings. Learning increases the probability of saying "grammatical" to grammatical strings and decreases it for nongrammatical strings. As people begin to learn, the probabilities start to covary with success, with a higher probability of correctly identifying strings that actually are grammatical. This means that the probabilities actually imply the epistemic status of the grammaticality judgement, ranging from a pure guess to reliable knowledge. The probabilities capture this information because without this correlation the system would not be successful and the relevant learning mechanism would not have evolved. However, the mechanism responsible for producing these probabilities need not explicitly represent that there is knowledge (i.e., that the representations induced by training and testing have the properties given in sect. 2.1.2). For example, there is no need for the mechanism to represent that there is something that is taken as reflecting the accuracy of the judgements, nor that the accuracy of the judgements is well founded in the learning history, nor that the self is the possessor of the knowledge.

Although participants' response probabilities suggest only a structure-implicit representation of the accuracy of their judgements, we do not know whether they have a

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

more explicit representation of it. One way to test whether they can represent the epistemic status of their judgements explicitly is to ask them to state their confidence in each classification decision (e.g., on a scale ranging from "guess," through degrees of being "somewhat confident," to "know"). If the confidence rating increases with the probability of responding correctly to each item, with random responding given a confidence of "guess," and deterministic responding given a confidence of "know," then the propositional attitudes implied by the probabilities have been used by the participant to explicitly represent the epistemic status of the grammaticality judgements; if confidence ratings are not so related to response probabilities, then epistemic status has been represented only implicitly.

The above tests only whether participants represent their ability to make judgements as knowledge. It is possible, as in the natural language case, that they know when they have the knowledge for judging grammaticality and when they are guessing, but still their knowledge of grammar rules is not represented as knowledge. This could be tested if we knew the actual content of participants' grammar rules. If the rules have been induced over time by some kind of optimal learning rule, then the epistemic status of the rules must be greater than just guesses. If participants, despite stating rules freely, or endorsing presented rules, nevertheless believe they are just guessing, then the rules have not been appropriately represented as knowledge. Also, if the rules had not been represented as knowledge, they may not be offered as descriptions of the grammar, because participants would not know that they knew anything. Of course, failure to state the rules in free report could also arise for other methodological reasons owing to the normal failings of free recall.

Establishing whether participants represent knowing their grammaticality judgements or grammar rules explicitly or implicitly is methodologically easier; the relevant research to date has focused on judgements. As noted above, one way to determine whether participants explicitly represent their ability to make judgements as knowledge would be to determine for each test item the probability with which it is given the correct response. If a plot of confidence against probability is a monotonically increasing line going through guess (0.5) to know (1.0), participants have fully used the implications of the source of their response probabilities to infer an explicit representation of their state of knowledge. If the line is horizontal, their knowledge is represented purely implicitly. If the line has some slope but participants perform above chance when they believe they are guessing, then some of the knowledge is explicit *and* some of the knowledge is implicit.

In artificial grammar learning experiments, participants typically make one or two responses to each test item so it is not possible to plot the confidence-probability graph just described, but it is not strictly necessary to do so. Consider the case where the participant makes just one response to each test item. We divide the items into those for which the participant makes a correct decision ("correct items") and those for which the participant makes an incorrect decision ("incorrect items"). If accuracy is correlated with confidence, the correct items should be a selective sample of those given a higher average confidence rating than the incorrect items. Conversely, if participants do not assign greater confidence to correct than incorrect items, then that is evidence that the slope of the graph is zero; that is, they do not represent their state of knowledge of their ability to judge correctly. If participants give a greater confidence rating to correct than incorrect items, that is evidence of at least some explicitness. If in this case, participants perform above chance when they believe they are literally guessing, that is evidence of some implicitness in addition to the explicitness.

Note that the previous paragraph presumes (1) a certain theory of how participants apply their knowledge (probabilistically, rather than deterministically) and (2) that the knowledge is largely valid. Reber (1989) has consistently argued that people's incidentally acquired knowledge of artificial grammars is almost entirely veridical. If people had applied partially valid rules deterministically, there would be no difference between confidence in correct and incorrect decisions, irrespective of whether the knowledge was attitude explicit. Thus, applying the procedure in different domains requires carefully considering how knowledge is applied in each.

Chan (1992) was the first to test whether participants explicitly represented knowing their grammaticality judgements. Chan initially asked one group of participants (the incidentally trained participants) to memorize a set of grammatical examples. In a subsequent test phase, participants gave a confidence rating for their accuracy after each classification decision. They were just as confident in their incorrect decisions as they were in their correct decisions, providing evidence that knowing was represented only implicitly. He asked another group of participants (the intentionally trained participants) to search for rules in the training phase. For these participants, confidence was strongly related to accuracy in the test phase, indicating that intentionally rather than incidentally trained participants represented their knowing more explicitly. Manza and Reber (1997), using stimuli different from Chan's, found that confidence was reliably higher for correct than incorrect decisions for incidentally trained participants. On the other hand, Dienes et al. (1995) replicated the lack of correlation between confidence and accuracy, but only under some conditions: the correlation was low particularly when strings were longer than three letters and presented individually. Finally, Dienes and Altmann (1997) found that when participants transferred their knowledge to a different domain, their confidence was not related to their accuracy.

In summary, there are conditions under which participants represent knowing grammaticality implicitly on most judgements, but there is sometimes evidence of having an explicit attitude of knowing. Even in the latter case, there is usually evidence of implicit knowledge: Both Dienes et al. (1995) and Dienes and Altmann (1997) found that even when participants believed they were literally guessing, they were still classifying substantially above chance.

Dienes et al. (1995) provided evidence that this type of implicit knowledge was qualitatively different from knowledge about which the participants had some confidence. When they performed a secondary task (random number generation) during the test phase, the knowledge associated with "guess" responses was unimpaired, but the knowledge associated with confident responses was impaired (to a level below that of the knowledge associated with "guess" responses). That is, this criterion is not just another curious way of categorizing knowledge: it may separate knowledge in a way that corresponds to a real divide in nature.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

4.4.3. Summary. In summary, when participants learn artificial grammars, there is evidence that for at least some of the acquired knowledge, participants represent the grammar of which the knowledge is predicated and can thus exert intentional control over which body of knowledge to apply. This intentional control indicates, by our analysis in section 3.4, that the participants have conscious knowledge of some content predicated of that grammar – in particular, the content they use to choose the grammar. There is no need to suppose, however, that participants were conscious of any further aspect of their knowledge (e.g., what the rules of their induced grammar were). If, based on task instructions, participants form the representation "I am thinking that I should apply the first grammar I studied," they are conscious of their desire to apply the first grammar. If the knowledge pertaining to this grammar is represented predication-explicitly, the mental specification that that is the grammar they want to apply may be sufficient to ensure that it does apply, so the participant has volitional control because of the predication explicitness of the representations formed during learning. The representations of the knowledge about the grammar may not make explicit that the rules are facts, however, or that the knowledge is knowledge. In that case, participants may have volitional control but may regard their responses as guesses, an outcome found by Dienes et al. (1995). In several studies, there was evidence that participants did not explicitly represent knowing many of their grammaticality decisions, thus they were not conscious of this knowledge as knowledge. The reason for this is precisely that participants did not have conscious knowledge of their grammar rules and hence could not know that their grammaticality decisions were based on sound knowledge.

These comments illustrate how one can empirically tease apart whether or not the knowledge is predication implicit or attitude implicit. This allows future research to determine which aspects of knowledge are left implicit in the representations formed during different types of learning. Such research could address whether different types of implicitness correspond to qualitatively different learning systems. In addition, future research needs to address other implicit learning paradigms (see Dienes & Berry 1997 and Stadler & Frensch 1998 for detailed reviews of implicit learning generally.)

5. Conclusion

In this target article, the natural language meaning of the implicit-explicit distinction was applied to knowledge representations, with knowledge taken as an attitude held towards a proposition. A series of different ways in which knowledge could be implicit or explicit followed directly from the approach. The most important type of implicit knowledge consists of representations that merely reflect the properties of objects or events without predicating them of any particular entity. The clearest cases of explicit knowledge of a fact are representations of one's own attitude of knowing that fact. We argued that knowledge capable of such fully explicit representation provides the necessary and perhaps sufficient conditions for conscious knowledge. This is consistent with Kihlstrom et al.'s (1992) suggestion that it is bringing knowledge representations "into contact with" the representation of the self that makes consciousness possible, because that connection defines the self as an experiencing agent in possession of the knowledge. Kihlstrom et al. suggested that this connection to the self is lacking in implicit perception; we agree, and add that the lack may be even deeper: the perceptual knowledge may lack not only representation of the self, but even predication to a particular event (e.g., what happened a few seconds ago).

Our analysis also corresponds in places to some recent analyses by Cleeremans (1997) and Dulany (1991; 1997). According to Cleeremans (1997), "knowledge is implicit when it can influence processing without possessing in and of itself the properties that would enable it to be an object of representation" (p. 199). Knowledge can be an "object of representation" if participants can metarepresent their representation of the knowledge as having various properties; for example, if they can metarepresent it as accurate (or inaccurate), as judged to be true (or false or undecided), or as properly caused (or not). Thus, Cleeremans's criterion corresponds to one aspect of the distinction between attitude implicit and explicit; in particular, to whether the metarepresentation (0) (that "the representation of *Fb is a fact* is possessed by the system"[21]) given in section 2.1.2 is formed. If the content of a piece of knowledge, acquired by a reliable process, can be specified by the participant even as a guess, then it is not implicit according to Cleeremans's criterion. As we argued in the section on artificial grammar learning, behaviour may indicate that a grammatical decision has been taken to be accurate (by consistent responding), but the participant may judge the decision to be a guess. Thus, the attitude of knowing implied by the participants' behaviour has not been explicitly represented. The piece of knowledge "this string is grammatical" is unconscious as knowledge, but it is conscious as a guess because the participant can entertain higher-order thoughts about it ("I guessed that this string was grammatical"). A deeper form of implicitness occurs when one cannot even entertain a higher-order thought about the knowledge; this corresponds to Cleeremans's definition of implicit and to complete attitude implicitness in our terminology.

Cleeremans argues that connectionist networks are particularly suitable for producing implicit knowledge, an analysis that agrees with our own (see Dienes & Perner 1996). In a connectionist network, the only information available for further transmission through the system is the activation of units (by assumption, for a real connectionist network, not a simulated one). Thus, knowledge embedded in weights is simply not available to be represented as accurate or inaccurate knowledge; hence it naturally satisfies Cleeremans's definition of implicit. On the other hand, Cleeremans argues that in a symbol system representations appear to have at least the potential to be attitude explicit because the system that uses them could always decide whether or not it possesses them. Dulany (1997) makes a stronger claim. Like us, he describes consciousness as involving an agent (I) holding an attitude towards some content; but according to Dulany, propositional content is always conscious.

Our analysis makes a distinction between predication explicitness (which could be a symbolic representation "Fb") and, among other things, explicit representation of attitude; only the latter representation would produce consciousness of the content Fb. It may be true as a matter of empirical fact that any predication explicit representation also allows attitude explicitness; then Dulany's claim would be true.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

This is a bold empirical hypothesis, but our analysis makes clear that there is no a priori reason for believing it to be true – why should a representation formed, for example, for some local need by a part of our perceptual system *inevitably* allow attitude explicit representations? In section 4.1 we indicated that the predication explicitness of some types of (factuality implicit) perceptual knowledge is an open testable question.

Both Dulany (1991; 1997) and Jacoby (e.g., Jacoby et al. 1992) argued that implicit processes change subjective experience (see also Perruchet & Gallego 1997). In our analysis, predication implicit knowledge (i.e., maximally implicit knowledge) can change behaviour and we take it for granted that such behavioural change is accompanied by conscious experiences. In a subliminal perception experiment, for example, the activation of the word form "red" may lead to a "red" response on a forced choice objective test. This behaviour would be accompanied by the thought "red pops into mind," or something similar. But the perceptual event would not have been consciously experienced as a perceptual event; that would have required the representation "I am seeing the word red on the screen" (fully attitude explicit knowledge) to be produced directly by the act of seeing the word red on the screen. The predication implicit representation "red" might trigger inferential thoughts to the effect that "I must have seen the word red on the screen." These higher-order thoughts enable the participant to be conscious of the possiblity of having seen red, but those inferences do not constitute the conscious perception of red. So, like Jacoby, Dulany, and Perruchet, we do suppose that implicit knowledge is often accompanied by conscious experience; one must simply be clear about what it is that the person is conscious of. We do not claim, however, that all implicit knowledge leads to conscious experience. The perceptual system could consider various perceptual hypotheses (e.g., predication implicit features, concepts, or schemata) before settling on one (e.g., Marcel 1983b), predicating it to an individual. The other hypotheses might never influence conscious experience at all (although they had the potential). Also, a representation may not itself lead to conscious experience, but it might cause other representations downstream of processing that produce conscious experience.

Similarly, an attitude implicit rule may lead one to feel good about a particular part of an English sentence or other grammatical string; this is a conscious experience, but not of the rule. A participant implicitly learning an artficial grammar might induce the rule "T can follow M," without predicating it of a grammar, representing it as a fact, or representing an appropriate attitude towards it. Nonetheless, the knowledge may make the bigram "MT" look familiar, inducing a conscious experience that "MT looks natural." The participant might infer the further thought: "in this grammar, perhaps T can follow M." If this happens, the participant, by observing his own behaviour, has induced a piece of explicit knowledge that coexists with prior implicit knowledge. Within the participant's knowledge box is the unconscious representation "T can follow M," not predicated of any particular grammar or represented as a fact. In addition, there is in the knowledge box the conscious representation "I see that MT looks natural." Sometimes the unconscious and conscious representations will contradict each other, as in the experiment by Bridgeman (1991) reported in section 4.1.

Our analysis of the meaning of implicit is in itself neutral on the question of whether different systems are responsible for producing knowledge of different degrees of implicitness. However, different degrees of implicitness will be useful for different purposes, and our view of the evidence is that different systems often do realize different degrees of implicitness in their knowledge (e.g., see sect. 4.1). Dienes and Berry (1997) reviewed the field of implicit learning and concluded that there was a natural divide between learning that produced knowledge about which participants did or did not explicitly represent the attitude of knowing (as we indicated in sect. 4.4 on artificial grammar learning). Dienes and Berry recommended picking out attitude implicit knowledge by using confidence ratings, looking at whether participants performed above chance when they claimed they were just guessing. This "guessing criterion" was found to be useful in separating types of knowledge that were qualitatively different in other respects (e.g., guessing knowledge was found to be resistant to secondary tasks as compared to knowledge about which participants had confidence); but it is still a testable empirical question whether it is attitude implicitness/explicitness that distinguishes different learning systems. We suggest that implicit learning is a type of learning resulting in knowledge which is not labelled as knowledge by the act of learning itself. Implicit learning is associative learning of the sort carried out by first-order connectionist networks (Clark & Karmiloff-Smith 1993; Cleeremans 1997; Dienes & Perner 1996; Shanks 1995). Explicit learning is carried out by mechanisms that label the knowledge as knowledge by the very act of inducing it; a prototypical type of explicit learning is hypothesis testing. To test and confirm a hypothesis is to realize why it is knowledge. Participants in an implicit learning experiment are quite capable of analyzing their responses and experiences, drawing inferences about what knowledge they must have. These explicit learning mechanisms, when applied to implicit knowledge, can lead to the induction of explicit knowledge. As a result, the guessing criterion is an imperfect (but still informative) guide for picking out implicit knowledge; it is not the guessing criterion but the nature of the underlying representations that defines the knowledge as implicit.

In summary, we have presented a framework that makes clear the precise ways in which knowledge can be made implicit. It indicates *why and how* various notions such as consciousness, verbalizability, and volition are related to each other and to the notion of explicit knowledge. It also suggests testable predictions about cognitive development, vision, learning, and memory.

ACKNOWLEDGMENTS

We wish to thank Bruce Bridgeman, John Campbell, R. Carlson, Peter Carruthers, Ron Chrisley, Greg Currie, Martin Davies, Tony Marcel, Shawn Nichols, and Gabriel Segal for invaluable discussions, and Peter Carruthers, John Kihlstrom, Pierre Perruchet, and Carol Seger for their informative reviews.

NOTES

1. This requires that there be a system that can go into at least two states, one state for the fact and another either for the negation of the fact or for staying noncommittal about the fact.

2. There is no provision in this system for being in one state to indicate this is knowledge and being in another state either to leave it open whether this is knowledge or to indicate that it is not knowledge.

3. As a point of interest, one should mention that what remain implicit in this case are *unarticulated* constituents of what is known (Perry 1986) in the sense that they do not find expression in the representational vehicle. As a result, the knowledge remains "situated" within the causal context of knowledge formation, and inferences drawn from this knowledge are valid only as long as this context is maintained (Barwise 1987; Fodor 1987a).

4. "Metarepresentational" here is used in the looser sense of modifying representational status (as used by Leslie 1987) and not in its usual strong meaning of representing the representational relationship (Pylyshyn 1978) as Perner (1991) has pointed out.

5. Representation of the truth of Fb does not replace the functional role of the knowledge box of mentally asserting Fb, a problem Frege grappled within his "Begriffsschrift" (see Currie 1982, Ch. 4). But it allows representation of false propositions within one's knowledge box without their becoming asserted. That is, by representing "Fb is not a fact" in the functional role of knowledge, Fb is represented but not asserted. What is asserted is that Fb is not a fact.

6. Perner (1991) reviews evidence that these abilities – pretend play, understanding temporal change, and understanding representations – emerge at about the same age of 18 months.

7. We are grateful to Peter Carruthers for having pointed out in response to an earlier draft that without this addition "through a generally reliable process" our criterion (3) and with it our definition of knowledge becomes otiose. The practical point of criterion (3) is to distinguish reliable from unreliable sources, but even the most reliable source can in principle fail. If one requires that process to be so reliable that it necessarily follows that it produces true representations, then criterion (3) would imply criterion (1), but at the cost of a practically useless criterion (3).

8. This self-explicitness can be applied separately to the four different aspects of knowledge:

(0s) I have R.
(1s) I have R which accurately reflects the fact that Fb.
(2s) I take (judge) R as accurately reflecting the fact that Fb.
(3s) I have R which has been properly caused by its content through a generally reliable process, e.g., I saw the fact Fb.

The following implications hold between these three types of self-explicitness for a rational agent who takes himself to be rational: (1s), (2s), and (3s) each imply (0s). (2s) implies (1s) because representing oneself as believing Fb implies that one represents Fb as true. In other words, one cannot represent oneself as believing something that one represents as false. Conversely, (1s) implies (2s) because if one represents R as true one should treat it as true. (3s) strongly suggests but does not strictly imply (1s) (and hence (2s)), since representing that the knowledge was properly caused implies that it ought to be accurate (i.e., that I should take it to be accurate).

9. Conditions (0), (i), (ii), and (iii) capture the everyday use of the word "know." Cognitive scientists generally use a broader definition, namely, requiring only conditions (0), (ii), and (iii) to hold; simply being false is not sufficient reason to prevent a piece of knowledge from being knowledge (e.g., Newton's Laws). Removing conditions (i) and (1) would not alter any of the conclusions that follow; note that (1s) given in Note 6 should still be included, as it follows from (2s), so our characterization of fully explicit knowledge stands as is.

10. For example Dretske (1995) speaks of being "conscious" or "aware" when we have information about something and represent it as such as shown by the appropriateness of our behaviour. In this usage what we have in mind needs to be expressed as being "consciously aware" to distinguish it from being "unconsciously aware" (which some might find a strange combination, because "aware" or "conscious" carries the connotation of being consciously aware).

11. Block (e.g., 1994; 1995) emphasises the subjective feel of conscious experiences (phenomenal consciousness) as central to the mystery of consciousness. Our concern and that of most cognitive sciences would be merely a case of "access consciousness" or "monitoring consciousness." There are, however, some interesting arguments to the effect that second-order mental states are necessary and sufficient for subjective feel (e.g., Carruthers1992; 1996).

12. This is exactly what a blindsighted person will say, when performing at random. The critical trick that Weiskrantz et al. (1974) used to get more convincing performance than Pöppel et al. (1973) did was to instruct the patient to guess: "I'll show you a light that you won't be able to see. Even though you can't see it, give it a guess and point to it" (Weiskrantz 1988, p. 187).

13. More technically expressed, the issue was whether one should represent the knowledge that every man is mortal as (1) a declarative axiom "$(\forall x\ \text{Human}(x) \supset \text{Mortal}(x))$" and then apply the general inference procedure "$[\forall x\ (F(x) \supset G(x))\ \text{and}\ F(b)] \Rightarrow G(b)$" which means roughly: If in the database you find for Variables F, G, x, and b the expressions "$\forall x\ (F(x) \supset G(x))$" and "$F(b)$" then add "$G(b)$" to the database, or (2) should one encode the relevant knowledge directly in a specialised procedure: "Human(b) (Mortal(b)." Our interest is in the difference between representing the regularity that being human implies being mortal either by means of the declarative implication sign "$\supset$" or by means of an inference procedure (production) symbolised as "$\Rightarrow$."

14. It might appear that learning systems, which are based on purely procedural knowledge, can make this evaluation on the grounds of negative feedback. The critical difference is that negative feedback in learning leads to a weakening of the response tendency for future inferences but it leaves the already made inference uncontested.

15. Another source of inferential limitations that makes for modularity is implicitness of property structure. If there is an inference from "male" to "shaves in the morning," it cannot be used on bachelors unless their being male is represented explicitly. So if one domain does not use the same property structure as another, even though their concepts overlap the two domains are modular with respect to one another.

16. Although Jacoby's method constitutes a clear methodological improvement, one must point out a remaining weakness. There is no guarantee that all participants will use the same criterion for excluding information. Consider: Is knowledge that makes predication explicit but leaves factuality implicit (e.g., "the word 'butter' being on the list") sufficient for exclusion? Probably not; it needs to be represented as a fact. But is even that sufficient? Consider the possibility that the origin of this piece of knowledge is not explicitly represented and that consequently no justification for one's judgement can be given; then people under justification pressure, unsure of their intellectual competence, might not consider it a reliable fact and not bring it under the exclusion criterion. In sum, although Jacoby's procedure undoubtedly provides a methodological advance in dissociating implicit from explicit memory, it still suffers from the ambiguities inherent in indirect and direct tests as measures of implicit and explicit knowledge. We will briefly return to the issue of resolving such ambiguity in our discussion of intentional control of knowledge of artificial grammars.

17. To claim that visually guided action can be based on predication implicit representation may be too radical as Evans (1975) has shown how limited linguistic communication would be without predication. However, visual perception of and action in one's immediate surroundings may be different because relations within one's egocentric space are much more constrained than relations between linguistically communicating partners. In Campbell's (1993) words, this is possible because the features can be used in a causally indexical way which linguistic communication cannot exploit to the same degree because people typically do not stand in exactly the same causal relation to what they communicate about.

18. There may be relevant evidence from subliminal perception for which unconscious perception of the meaning of single words is possible but the subliminal perception of the meaning of word combinations is difficult to demonstrate (Greenwald 1992;

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

Kihlstrom 1996) – perhaps because the interpretation of combinations requires explicit predication.

19. Dokic (1997) pointed out that the above formulation of the memory trace still leaves room for counterexamples. In order to ensure a true episodic memory the encoding has to be self-referential in Searle's (1983) sense: "I know that (`butter' was on the list and this knowledge comes directly from my past experience of the list)." The parentheses are added to bring out more sharply the syntactic embedding that makes "this knowledge" self-referential.

20. For this reason one can speak of association but not of inference. Inferences go from state of affairs to state of affairs, that is, reasoning of the form "whenever X is the case then Y must be the case." But that means X and Y are predicated of particular occasions. That associative processes but not inferences are possible implicitly and without consciousness is reminiscent of Sloman's (1996) suggestion that implicit knowledge is tied to associative processes and explicit knowledge to rule governed inference processes.

21. We previously called this distinction "content implicit vs. explicit" (Dienes & Perner 1996).

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

The developmental progression from implicit to explicit knowledge: A computational approach

Martha Wagner Alibali[a] and Kenneth R. Koedinger[b]

[a]*Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213;* [b]*Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213.* **alibali@andrew.cmu.edu www.psy.cmu:edu/psy/faculty/malibali.html koedinger@cmu.edu act.psy.cmu.edu/ACT/people/koedinger.html**

Abstract: Dienes & Perner (D&P) argue that nondeclarative knowledge can take multiple forms. We provide empirical support for this from two related lines of research about the development of mathematical reasoning. We then describe how different forms of procedural and declarative knowledge can be effectively modeled in Anderson's ACT-R theory, contrasting this computational approach with D&P's logical approach. The computational approach suggests that the commonly observed developmental progression from more implicit to more explicit knowledge can be viewed as a consequence of accumulating and strengthening mental representations.

Dienes & Perner (D&P) consider the procedural-declarative distinction in light of their theory of implicit and explicit knowledge, arguing that nondeclarative knowledge can take multiple forms. We provide empirical support for this view by describing two lines of research about the development of mathematical reasoning. Both lines of work demonstrate how knowledge becomes increasingly explicit with development (cf. Karmiloff-Smith 1992).

One line of research focuses on students' ability to infer mathematical functions from tables of x,y pairs, for example {(2,5) (3,7) (4,9)} (Haverty et al. 1999; Koedinger et al. 1997). Some students are able to generalize the pattern to both small and large values of x (e.g., 5 and 93), and to articulate the pattern in a general rule, either in verbal form (e.g., "to get y, double x and add 1") or symbolic form (e.g., $y = 2x + 1$). Other students are unable to articulate the pattern, but are able to find y for both small and large values of x. Such students appear to have discovered the general rule, and can apply it across multiple instances of x, despite being unable to state the rule. It is important to note that individual patterns of skill suggest a developmental progression, such that students who are initially able to find y only for small values of x, later learn to find y for large values of x, and ultimately learn to articulate the pattern, first in words and then in symbols.

A second line of research focuses on 8- to 10-year-old children's ability to solve and explain equations of the form $3 + 4 + 5 = \_\_ + 5$ (Alibali 1999; Perry et al. 1988). Some children are able to articulate their procedures for solving such problems in both speech and gestures. Other children express at least some of their procedures for solving the problems only in gestures, and not in speech. However, the procedures these children express uniquely in gestures can be accessed in a rating task (Garber et al. 1998). Specifically, children give higher ratings to solutions derived from procedures they express in gestures than to procedures they do not express at all. Moreover, in the equation task, as in the function-finding task, more implicit knowledge appears to precede more explicit knowledge in development (Goldin-Meadow et al. 1993).

Thus, students are sometimes able to apply knowledge that they are unable to articulate, and sometimes able to express knowledge in gestures that they cannot express in speech. These examples underscore the notion, which D&P emphasize, that a simple dichotomy between implicit and explicit knowledge is inaccurate. However, it is not entirely clear how D&P's theory applies to these tasks. In their terms, it seems likely that knowledge that can be applied but not articulated, or gestured but not spoken, would be characterized as factuality-implicit but predication-explicit. As they put it: "One can think of the gestures as . . . sign[s] of thoughts about reality that have not yet been recognised as being about reality (implicit factuality)" (sect. 4.3). The implications of D&P's logic-based theory, however, are less clear for procedural knowledge of tasks like equation-solving and function-finding than for declarative knowledge like "this is a cat" (sect. 2.1.1). It is not clear, for example, what leverage we get by thinking about procedural knowledge (e.g., how to find a function or solve an equation) as a "fact."

A computational theory, such as ACT-R (Anderson & Lebiere 1998), may shed additional light on the implicit-explicit knowledge distinction in procedural tasks. In ACT-R, implicit procedures are modeled by production rules, which set goals and perform actions. The knowledge contained in production rules is not directly accessible to other production rules; hence it cannot be used in other forms of reasoning. One consequence of this is that the knowledge contained in production rules cannot be verbalized.

For procedures to be verbalized, elements of them must be encoded in declarative memory, *and* those declarative elements, or "chunks," must be accessible to other production rules that have the specific purpose of verbalizing thoughts (language productions). In ACT-R, declarative chunks have an associated activation level, and activation must be high for chunks to be accessible to language productions. Chunks that are weakly activated, like tentative hypotheses about what equation components or quantitative relations are relevant, may fail to fire complex language productions. They may be sufficient, however, to fire simpler, better-practiced productions for generating gestures.

This ACT-R interpretation provides insight into why the developmental progression from implicit to explicit knowledge is often observed in development (e.g., Clements & Perner 1994; Karmiloff-Smith 1986). In both the ACT-R theory and D&P's theory, explicit knowledge representations can be viewed as elaborate versions of implicit knowledge representations. In ACT-R, learners' knowledge becomes more explicit when they strengthen declarative chunks or when they acquire new language productions that allow them to express ideas they could not express before. Thus, explicit forms of knowing require not only the declarative chunks themselves, but also language productions to express that

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

knowledge. In D&P's theory, learners' knowledge becomes more explicit when they add "I know that" (attitude) to "it is true that" (factuality) to "this" (individual) to "is a" (predication) to "cat" (property).

As a computational theory, ACT-R leads to clear behavioral predictions, which may be at odds with or, at least, are not easily interpretable within D&P's theory. For example, ACT-R predicts that someone can have explicit declarative knowledge (e.g., can state the Pythagorean theorem) but may lack the corresponding implicit procedural knowledge (e.g., cannot apply it in context). Computational theories like ACT-R or neural networks also allow for distinctions between implicit and explicit knowledge that do not involve the addition of new knowledge. In such theories, knowledge that was once inaccessible to other processes, such as verbalization, increases its probability of being accessed as its strength increases through use.

D&P have made it clear that careful analysis of knowledge, in its multiple forms, is critical to advancing our understanding of the brain and behavior. Such insights are not only interesting scientifically, but particularly important to the everyday world of educational decision making. To design effective instructional methods and assessment techniques, it is essential to understand implicit knowledge and its role in performance and development (Koedinger & MacLaren 1997; Koedinger et al. 1997).

ACKNOWLEDGMENT

Work described in this commentary was supported by a grant from the James S. McDonnell Foundation.

Individuals, properties, and the explicitness hierarchy

Alex Barber

Department of Philosophy, University of Sheffield, Sheffield S10 2TN, United Kingdom. **a.barber@sheffield.ac.uk**

Abstract: The scenario used by Dienes & Perner to show that individual representation can be implicit when property representation is explicit can be adapted to show that property representation can be implicit when individual representation is explicit. So there is no hierarchy of explicitness, contrary to their claim. There is a reading of the "implicit/explicit" distinction that does appear to exhibit an asymmetry parallel to that alleged to hold between individual and property. But this is not a distinction Dienes & Perner mention, nor is it one that could be easily incorporated into their framework.

A useful feature of Dienes & Perner's (D&P's) discussion is that their characterization of the implicit/explicit distinction makes room for the possibility of various permutations in how elements in any given episode of knowledge can be assigned to the different sides of the divide. D&P go on, however, to claim that only certain of the logically possible permutations are genuinely possible (Fig. 1). Specific aspects of this further claim are unconvincing.

By adapting Strawson's (1959) naming game, D&P allow the possibility of implicit individual-representation combined with explicit property-representation (sect. 2.1.1). Their reasoning seems to be that the player hypothesized as performing this game can afford not to represent the individual explicitly – that is, the player has no need of a capacity to go into a this-object/some-distinct-object state (n. 1 and 2) – because there is only one salient object for the property to be inhering in. But a variant game is available to establish a possibility that D&P rule out: explicit individual-representation combined with situated, implicit property-representation. The player is called on to verbally identify a highlighted personage (say, a figure under a spotlight). The word "JFK" conveys the information that JFK has the property of being highlighted. Yet what is made explicit within the vocabulary of this game are only the identities of JFK, Madonna, and so forth. Because the player manifestly knows that it is the property of being highlighted that is predicated of the particular individual, this component of the knowledge must be implicit. It need not be made explicit because it is the only salient property, just as for the original version of the game there was only one salient object.

Judging by their comments on a similar game (sect. 2.1.1, para. 7), D&P would most likely reply that the *knowledge underlying* successful performance in this second game must involve explicitness for both components, individual and property. Players must be able to go into highlighted/not-highlighted state for the individual to decide whether that individual is being highlighted. But could equivalent reasoning not be deployed against their own original use of the naming game? Players there must be able to go into a "the-object-in-front-of-me/not-the-object-in-front-of-me" state for each property, to refrain from calling out "dog" on recognition of the dogginess of some distinct object – an imagined dog from the players' childhood, or a dog in the periphery of their vision.

A second reply is potentially available to D&P. They could claim that the content of the relevant knowledge in the variant naming game is that *the object under the spotlight has the property of being JFK.* In this case, it would still be the object-representation that is implicit, the property-representation that is explicit. But inasmuch as *every* proposition can be re-expressed in this way (see Schiffer 1987, p. 51, on pleonastic properties), the claim to have found an asymmetry would be unprincipled.

In the remainder of this commentary I argue that there is a reading of the "implicit/explicit" distinction that *does* appear to exhibit an asymmetry parallel to that alleged to hold between individual and property. But this is not a distinction D&P mention, nor is it one that could be easily incorporated into their framework.

Contrast two sentence-couples:

> Noah believed the lion in front of him was hungry. He ran away fast.

> Noah believed the lion in front of him was hungry. Not realising it was a lion, he attempted to feed it some spinach leaves.

The noun phrase (NP) in the complement clause expressive of the allegedly believed proposition ("the lion in front of him") is the same in each case, but clearly there is also some difference between the two attributions. In traditional terminology the first is a *de dicto* attribution, the second, a *de re* attribution. Following Quine (1956), we might rephrase the second by extracting the NP thus;

> Noah believed, of the lion, that it was hungry

so as to avoid any implication that Noah had, as it is often put, conceptualized the lion *as* a lion. No such implication would need to be avoided in the first attribution. This is no place to stake out a position on the status of the *de re/de dicto* distinction, so I limit myself to four brief points.

First, although it is usually discussed in relation to belief attribution, this contrast is equally manifest in knowledge attribution.

Second, the utility of NP extraction does not appear to carry over to the verb phrase (VP) ("was hungry"). There are no easily imaginable situations in which the following would effect a useful implication restriction:

> Noah believed, of hungriness, that the lion in front of him was thus.

So there is an asymmetry that matches the asymmetry claimed by D&P between individual and property, and which serves, perhaps, as a tacit source of that claim.

Third, the *de re/de dicto* distinction is plausible as a potential landing-pad for the "implicit/explicit" label, particularly in light of the etymological connection with language (*cf. Introduction* and sect. 3.2). (On the other hand, one reason for thinking that cognitive psychologists may aspire to have no professional interest in *de re* attributions is this: The extent to which an attribution is *de re*

corresponds to the degree of absence of any "cognitive oomph" – witness the spinach leaves.)

Fourth, if the "implicit/explicit" label were tied down to the *de re/de dicto* distinction, or if the latter distinction were the source of the alleged asymmetry between individual and property, then it would be worth taking note of the following fact (given the import of sect. 3.1 of the target article): The cognitive episode described as Noah's believing of the lion that it is hungry need not be an unconscious cognitive episode.

I close with a historical observation, potentially relevant to but certainly not at odds with the perspective of D&P. At varying points throughout this century, good evidence for the existence of knowledge (or belief, desires, etc.) has come into conflict with more traditional and conservative conceptions of knowledge. Such conflict generated reservations about particular attributions of knowledge, reservations that were often acknowledged by qualifying "knowledge" with prefixes such as "unconscious," "unverbalizable," "nonconceptual," "dispositional," "involuntary," "nonpromiscuous," or "externalist," or the suffix "-how," according to the absent "essential" feature of "genuine" knowledge. The prefix "implicit," like "tacit," can be thought of as functioning as a rubric for these terms. Understanding the label in this way, there is no *a priori* reason why all the subgenres of implicit knowledge should relate to one another in any more robust a fashion than that each is correlated with one shortcoming or another of the traditional conception of knowledge. So it would be neither surprising *nor especially objectionable* if there are understandings of "implicit/explicit" that fail to gel with that offered by D&P.

Volitional control in the learning of artificial grammars

Peter A. Bibby and Geoffrey Underwood

School of Psychology, University of Nottingham, Nottingham NG7 2RD, United Kingdom. **{pal; gju}@psychology.nottingham.ac.uk**

Abstract: Dienes & Perner argue that volitional control in artificial grammar learning is best understood in terms of the distinction between implicit and explicit knowledge representations. We maintain that direct, explicit access to knowledge organised in a hierarchy of implicitness/explicitness is neither necessary nor sufficient to explain volitional control. People can invoke volitional control when their knowledge is implicit, as in the case of artificial grammar learning, and they can invoke volitional control when any part of their knowledge representation is implicit, as can be seen by examining "feeling of knowing" phenomena.

Dienes & Perner (D&P) argue that *volitional control* in artificial grammar learning (Dienes et al. 1995) is best understood in terms of the distinction between implicit and explicit knowledge representations. Assuming that direct control of one's knowledge requires that knowledge of a grammar has been explicitly labelled as knowledge that can be used in a particular situation, then D&P's argument hinges on the difference between two statements about a grammar. First, "M can be followed by T," which does not include any specific information about a grammar; second, "g has the property that 'M can be followed by T,'" which denotes a specific grammar and makes explicit the predication of "M can be followed by T" to the grammar g. If participants have only implicit knowledge of the first type, they should not be able to invoke volitional control. This claim raises two issues. First, can the Dienes et al. (1995) study be explained without the distinction between implicit and explicit knowledge? Second, given D&P's hierarchy of explicitness, is it possible to invoke volitional control when the subordinate level of the hierarchy is implicit and the super-ordinate is explicit?

D&P's explanation of Dienes et al.'s (1995) experiments relies on the fact that the two grammars are discriminable. A simple simulation, run here specifically to test their interpretation, demonstrates that the two grammars are easily discriminated, provided that subjects can pick up the covariance in the letter bigrams that are used to construct the letter strings. In the simulation, which was run 10 times, 2 randomly sampled training sets were generated on the basis of Dienes et al.'s (1995) grammars. For both grammars the initial and final bigrams were the same. The training sets consisted of 20 strings between 5 and 9 letters in length. There were 2 test sets of 20 items, generated, 1 for each grammar, such that none of the training items were re-presented. A measure of item-grammar concordance for the test sets was calculated. For example, the test string VTVM is comprised of three bigrams, VT, TV, and VM. If VT appeared 9 times, TV appeared 6 times, and VM appeared 5 times in the training set for grammar 1, then the letter string VTVM received a cumulative frequency score of 20 for grammar 1. If VT appeared three times, and neither TV nor VM appeared in the training set for grammar two, then the test string VTVM received a cumulative frequency score of three for grammar two. For each run of the simulation an average cumulative frequency score was generated for all the items in the two test sets. The results of the simulation are shown in Figure 1.

The two training sets are not only discriminable, they are differentially discriminable. This second-order effect may be precisely the kind of information participants use to decide that a test item belongs to a specific, previously seen set of items. That the two grammars are differentially discriminable undermines D&P's claim that the volitional control of grammar choice cannot be done by means of a compound property. They argue that if compound properties are used to distinguish between grammars, then speed of identification should predict classification, and it does not. Assuming that the degree of item-grammar concordance predicts the time taken to assess whether an item belongs to a particular grammar (as would be expected were this implemented in a neural network simulation) then, as Figure 1 shows, the time taken on average to identify an item as belonging to grammar one and to then assess whether it belongs to grammar two, and vice versa, would be approximately equal. This suggests that there is no reason why speed of identification should predict classification, provided that subjects assess the plausibility of an item as belonging to both grammars. We argue that the differential discriminability of the items can be used to identify an item as belonging to a grammar, decide which grammar it belongs to, and explain why speed of identification does not predict the success of the classification process. Accordingly, there is no need for the knowledge used to invoke volitional control in this task to be explicit.

The second issue is whether there is a need for a hierarchy of explicitness to explain volitional control. There is evidence to suggest that people can represent the knowledge that "X has the property Y" without knowing what Y is. More precisely, knowledge at

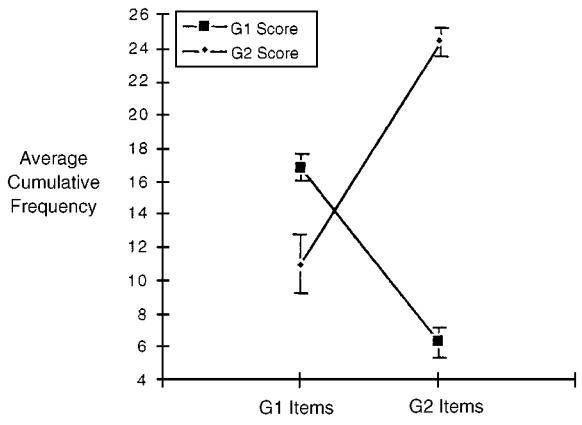

Figure 1 (Bibby & Underwood). The average cumulative frequency scores per simulation run for test items generated by both grammars (G1 and G2) and scored for both grammars.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

the bottom of D&P's hierarchy can be implicit whilst knowledge at the top of the hierarchy is explicit. The "feeling of knowing" phenomena depends on someone being able to say "I know that X has the property Y but I don't know what Y is." For example, X may be the addition of two numbers and Y, the answer to the sum. People can make the judgement that they know what the answer is but cannot remember it, or they know what the answer is and they could remember it given time, or they know what the answer is and this is it. The degree of certainty that people have about their "feeling of knowing" predicts whether they will spend time trying to retrieve the answer or spend time recalculating the answer (Reder 1988). People use volitional control to decide which strategy to adopt without direct access to their knowledge. Indeed, in the case of "feeling of knowing," it is a necessary condition that knowledge at the bottom of the hierarchy is implicit and that it is explicitly known to be implicit.

Direct, explicit access to knowledge organised in a hierarchy of explicitness is neither necessary nor sufficient to explain volitional control. People can invoke volitional control when their knowledge is completely implicit and they can invoke volitional control when any part of their knowledge representation is implicit.

Unconscious motivation and phenomenal knowledge: Toward a comprehensive theory of implicit mental states

Robert F. Bornstein

Department of Psychology, Gettysburg College, Gettysburg, PA 17325.
bbornste@gettysburg.edu

Abstract: A comprehensive theory of implicit and explicit knowledge must explain phenomenal knowledge (e.g., knowledge regarding one's affective and motivational states), as well as propositional (i.e., "fact"-based) knowledge. Findings from several research areas (i.e., the subliminal mere exposure effect, artificial grammar learning, implicit and self-attributed dependency needs) are used to illustrate the importance of both phenomenal and propositional knowledge for a unified theory of implicit and explicit mental states.

Research examining the contrasting dynamics of implicit and explicit mental processes can provide great insight into a variety of psychological issues. Dienes & Perner's (D&P's) analysis of implicit and explicit knowledge makes a valuable contribution to this effort, but their arguments apply primarily to those domains of human mental life that have traditionally been studied by cognitive scientists (e.g., perception, memory, learning, language use). This commentary argues that a comprehensive theory of implicit and explicit knowledge must incorporate the motivational domain, as well.

Studies have demonstrated that human motives can be usefully divided into two broad categories (McClelland et al. 1989). *Explicit motives* (also called *self-attributed motives*) are accessible to conscious awareness, and can be accessed via direct verbal report. *Implicit motives,* on the other hand, affect behavior indirectly, are inaccessible (or only partially accessible) to conscious awareness, and are not acknowledged directly by the actor. As these definitions illustrate, there are some noteworthy parallels between the implicit-explicit motive distinction and the implicit-explicit distinction discussed by D&P in their analysis of perception, memory, learning, and language use.

To extend the implicit-explicit knowledge distinction to the motivational domain, it is necessary to expand D&P's definition of knowledge, which emphasizes mental representations of "facts" (i.e., verifiable propositions) regarding oneself and external objects. A more complete and inclusive definition of knowledge must account for feeling states as well, along with the motivational representations derived from them. In other words, a comprehensive theory of implicit and explicit knowledge must include *phenomenal knowledge* (e.g., knowledge regarding one's internal affective and motivational states), as well as *propositional knowledge* (i.e., "fact"-based knowledge).

A logical consequence of D&P's conceptual framework is the hypothesis that implicit propositional knowledge may sometimes be expressed phenomenologically (see also Schacter 1987). This is certainly the case in studies of the subliminal mere exposure (SME) effect (Bornstein 1992). Numerous experiments have shown that visual stimuli presented under degraded conditions result in implicit memory traces for those stimuli: Participants cannot report having seen the stimuli before (propositional), but they nonetheless describe these stimuli as more pleasing or likeable than similar unfamiliar stimuli (phenomenal). Parallel findings have emerged in those studies of artificial grammar learning (AGL), which require participants to judge the likeability or pleasantness of novel grammatical and ungrammatical letter strings (Manza & Bornstein 1995).

In this context, it is important to note that just as certain experimental manipulations can dissociate implicit and explicit memory effects, experimental manipulations can help disentangle the effects of implicit and explicit motivational states. For example, manipulations that induce negative mood states temporarily alter participants' implicit – but not explicit – dependency strivings (Bornstein 1998). In contrast, manipulations that modify participants' beliefs regarding the acceptability of dependent (or autonomous) behavior affect explicit – but not implicit – dependency needs (Bornstein et al. 1994). There is some preliminary evidence that similar results are obtained in the achievement, intimacy, and power domains, as well (McClelland et al. 1989).

The study of implicit and explicit mental states has the potential to unify psychology, connecting ostensibly unrelated phenomena in diverse domains and topic areas (e.g., cognitive, developmental, social, personality). To develop the most heuristic conceptual framework possible, we must broaden that framework to include the entire range of implicit and explicit phenomena relevant to human mental life. Only then will a theory of implicit and explicit knowledge contribute to the unification of psychology rather than its continued fractionation.

Time and the implicit-explicit continuum

Jill Boucher

Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom. **j.boucher@warwick.ac.uk**

Abstract: Dienes & Perner's target article contains numerous but unsystematic references to the implicit or explicit knowledge of the temporal context of a known state of affairs such as may constitute the content of a propositional attitude. In this commentary, the forms of cognition that, according to D&P, require only implicit knowledge of time are contrasted with those for which explicit temporal knowledge is needed. It is suggested that the explicit representation of time may have been important in human evolution and that certain developmental disorders including autism may be (partly) caused by defective ability to represent time.

Dienes & Perner (D&P) mention time at numerous points in their target article, but they do not systematically describe the role of the representation of time in their analysis of knowledge explicitness. A search through their article suggests, however, that if their analysis of knowledge explicitness is correct (and I assume here that it is at least substantially correct), then the explicit representation of time is critical to a raft of higher cognitive functions. Specifically, D&P argue that:

1. Information concerning the time of occurrence of an event can only be left implicit for events occurring in the present, and representations of propositional content that do not mark time and factuality are only useful representing the here and now[1]

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

(sect. 2.1.1). Representations that do not explicitly represent time allow only for the following forms of cognition:

a. Having propositional content that can be entertained by the system (sect. 2.1.2).
b. Learning associations and event sequences, and having expectations as to what happens next (sect. 4.3); learning procedures (sect. 3.3).
c. Responding to indirect tests of knowledge (sect. 3.2); guessing (sect. 4.2).
d. Fodorian modular processing (sect. 3.3).
e. Contention scheduling and automatic action (sect. 3.4).
f. Visually guided movement (sect. 4.1).

2. Representations of time past are always explicit (sect. 2.1.1), and the explicit representation of time within propositional content makes the following forms of cognition possible:

a. (1) Having fully explicit propositional content (sect. 2.1.1). (2) Having explicit knowledge of propositional attitude (knowing about knowing); (sect. 2.1.3 and 3.1), including entertaining propositional content (R) as true (accurate), judging R to be true and therefore holding a belief, and knowing the causal origin of R (sect. 2.1.2). (3) Having explicit knowledge of oneself as the holder of a propositional attitude (sect. 2.1.3 and 3.1).
b. Understanding change over time (sect. 2.1.1); keeping track of past states of affairs and comparing them with the present state of affairs (sect. 4.1); explicit memory and recollection (sect. 2.1.1); declarative knowledge (sect. 3.3); hypothetical reasoning (sect. 3.3); pretense and understanding representation (sect. 2.1.1 and Perner 1991); planning (sect. 3.3); causal understanding (sect. 4.1); consciousness (sect. 3.0 and 3.1).
c. Responding to direct questions in laboratory tests of memory/knowledge (sect. 2.1.1 and 3.2); knowing and remembering (sect. 4.1).
d. Central processing (sect. 3.3).
e. SAS (supervisory attentional system) control of action; voluntary action.
f. Verbally guided (declarative responding (sect. 4.1).

D&P's claims concerning the role of explicit time representation in higher cognitive functions are consistent with the hypothesis that developments in the representation of time played a critical role in human evolution (Boucher 1998a). The forms of cognition listed under (1), above, correspond in general to the cognitive abilities of animals, whereas the forms of cognition listed under (2) include most of the cognitive abilities generally considered to be unique to humans. One would want to add the claim that explicit time representation is critically involved in language (cf. Elman 1990), an ability D&P assume under (2) above, but do not discuss.

D&P's claims are also consistent with the hypothesis (Boucher 1998b; in press a; in press b) that autism and developmental language disorders are wholly or partly caused by genetically determined deficits in the ability to process time. With regard to autism, the forms of cognition that can operate in the absence of the explicit representation of time correspond in general to the cognitive abilities that may be spared in autism, whereas the forms of cognition for which the explicit representation of time is necessary correspond to the cognitive impairments that characterise autism. D&P do not discuss language; their paper therefore has no direct implications for our understanding of child language disorders. I have argued, however, that child language disorders are caused by defective timing mechanisms related to but distinct from those that cause autism, and that this explains why autism and language disorders commonly occur together in individuals and within families (Boucher, in press b).

If what D&P say about time is correct and, as I suggest, consistent with my hypotheses concerning evolution and certain developmental disorders, then knowledge explicitness, human evolution, and these disorders are linked by a common dependence on species-specific time processing mechanisms. The nature of these mechanisms is unclear. However, recent research suggests that computations over the outputs of multiple oscillators, or so-called biological clocks, may have evolved to subserve time-dependent cognitive and linguistic abilities in humans, in addition to the perceptual and motor skills they primarily subserve in animals (Boucher 1998a; Brown & Vousden 1998; Gallistel 1990).

NOTE

1. D&P do not always distinguish clearly between time and factuality (constituents 4a and 4b of propositional content). Where they state that explicit factuality is necessary for x, without also stating that the explicit representation of time is necessary, I may have overinterpreted them as meaning that the explicit representation of both factuality and time are necessary for x.

Implicit and explicit representations of visual space

Bruce Bridgeman

Department of Psychology, University of California, Santa Cruz, CA 95064.
zzyx.ucsc.edu/psych/psych/faculty/bridgeman.html
bruceb@cats.ucsc.edu

Abstract: The visual system captures a unique contrast between implicit and explicit representation where the same event (location of a visible object) is coded in both ways in parallel. A method of differentiating the two representations is described using an illusion that affects only the explicit representation. Consistent with predictions, implicit information is available only from targets presently visible, but, surprisingly, a two-alternative decision does not disturb the implicit representation.

By differentiating several levels along an implicit-explicit dimension, Dienes & Perner (D&P) are able to integrate several previously unrelated or contradictory literatures into a consistent theoretical framework. Knowledge is represented at one or another level of explicitness, depending on its nature, the way in which it was acquired, and its function.

A particularly illuminating domain for studying questions of the relation of implicit to explicit knowledge is the representation of the location of an object in visual space. Here, knowledge about the same parameter, spatial location, is coded in two distinct neurological systems that differ in their degree of explicitness. Most other cases, in contrast, concern either an explicit or an implicit representation in isolation. Thus the contrast between the properties of explicit and implicit knowledge is particularly clear here, where questions of the content of the representations are not confounded with their function.

For these reasons, D&P devote a significant amount of space to this example. Their review of experiments on induced motion is referenced to Bridgeman (1991) and Bridgeman et al. (1997), but concerns experiments conducted earlier (Bridgeman et al. 1981). In those experiments we compared an explicit, cognitive representation and an implicit, sensorimotor representation in contrasting experimental conditions. One condition injected a motion signal into an explicit representation of visual space by using a moving frame, which induced an opposite motion into an objectively fixed visual target. Thus a motion signal was present in explicit knowledge. Subjects always pointed open-loop to this target in the same direction, however, regardless of whether it had just apparently shifted to the left or to the right, showing that the motion signal was not affecting the implicit knowledge that controls visually guided behavior.

In other trials of this experiment, subjects nulled the out-of-phase induced motion by adding an in-phase real motion to the target. Because this task was based on perceived motion it nulled the explicitly represented motion signal, so that the subjects could state, experience, and remember that only the background frame, and not the target, was moving. Nonetheless, when asked to point to the target, their behavior depended on the true egocentric po-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

sition of the target rather than on their experience. In the first case, position changed explicitly but not implicitly, a case of A and not B; in the second case, position changed only implicitly, a case of B and not A, thus defining a double dissociation.

There is a danger that motion and position might be confounded in this design, however. This is why we sought another method, one inducing an illusion of static position in the cognitive system without the target, the background or the subject ever moving during the stimulus exposure. The induced Roelofs effect (Bridgeman 1991) meets these criteria. A static frame is projected in an otherwise uniform field, either centered about the subject's midline, or offset to the left or right. If the frame is offset, a single static target within the frame will be perceived not in its true position, but deviated in a direction opposite the offset of the frame. Subjects are shown five possible positions of the target, and asked to indicate verbally which of the positions was presented.

We contrast this with an open-loop pointing measure, and here D&P's distinction between pointing as a communicative act (explicit) and as an instrumental act (implicit) becomes critical. At Perner's earlier suggestion (personal communication) we have eliminated this ambiguity by asking subjects to jab a lever, making a loud clacking sound, rather than pointing to the target. They do not communicate anything; rather, they simply do a job. Now all subjects show independence from frame position in their motor behavior, even though they experience the Roelofs effect as determined with the verbal measure.

D&P note that factuality can be left implicit only for the present. This property of implicit representations predicts that our sensorimotor representation, which is factually implicit, should operate only on visible stimuli, and should quickly decay in the absence of stimulation. Indeed, the duration of Roelofs-free jabbing after the offset of a stimulus can indicate the duration of the psychological present.

In our earlier experiments we indeed found a Roelofs effect for pointing if the action was delayed. Recently we have measured this phenomenon more closely, using the presence of a Roelofs effect in jabbing as an indication that spatial information is no longer available from the implicit sensorimotor system. After a delay of 1 sec, subjects retain the implicit spatial information, but at 2 sec they begin to show small but significant Roelofs effects. Thus the implicit system can hold spatial information for only 1–2 sec, in agreement with earlier estimates of the psychological present by subjective methods (Fraisse 1963).

Further investigating the implicit representation, we independently arrived at the prediction that a decision, characteristic of explicit processing, would force the reappearance of a Roelofs effect even in immediate jabbing in the presence of an offset frame. The result, published after D&P's paper was accepted, was a surprise – jabbing to one target was nearly identical to jabbing at one of two physically identical targets located 5° apart (Bridgeman & Huemer 1998). No Roelofs effect appeared in either condition, contradicting both our prediction and that of Dienes & Perner. The only difference between the conditions was a slight increase in variability when two targets were present. Our explanation is that the explicit cognitive system is somehow able to inform the sensorimotor system about which target to jab, and that system uses its own spatial information to find the target, regardless of frame position.

Nonconceptual content and the distinction between implicit and explicit knowledge

Ingar Brinck

Department of Philosophy, Lund University, S-222 22 Lund, Sweden.
ingar.brinck@fil.lu.se

Abstract: The notion of nonconceptual content in Dienes & Perner's theory is examined. A subject may be in a state with nonconceptual content without having the concepts that would be used to describe the state. Nonconceptual content does not seem to be a clear-cut case of either implicit or explicit knowledge. It underlies a kind of practical knowledge, which is not reducible to procedural knowledge, but is accessible to the subject and under voluntary control.

In this commentary I would like to point to some cases in which the knowledge involved does not seem to fit into Dienes & Perner's (D&P's) schema. This is primarily the kind of knowledge that lies behind practical competence. In some cases, at least, it cuts across D&P's categorisations.

D&P rely on the representational theory of mind (RTM, Fodor 1978) to describe knowledge representations, which means that explicit knowledge must be represented propositionally. RTM squares very well with their theory. But there seems to be knowledge that can neither be described in the framework of RTM nor does it fit into the schema of the implicit and the explicit. Whether RTM is correct or not is nevertheless more of a technical than a substantial question in the present context, and I will not discuss it further.

Let me instead turn to the notion of nonconceptual content, that is, content that is independent of concepts. A subject may be in a state with nonconceptual content without having the concepts that would be used to describe the state. It is evoked to explain behaviour that relies on representations, but cannot be captured by concepts.

Examples of nonconceptual content are the richness of perceptual experience that exceeds conceptual description, and infant and animal perception of the environment, the content of which diverges from conceptual descriptions of the environment. Nonconceptual content has correctness conditions, although it does not constitute propositional belief that can be assigned a truth-value. It presents things to the subject and can do so adequately or inadequately.

Nonconceptual representation of categories will be context-sensitive and influenced by the properties of the subject, the ongoing interaction between subject and environment, and other factors that emerge in the context. It does not involve general conceptual identification or metarepresentations of relations.

What is the place of nonconceptual content in D&P's theory? It cannot constitute explicit, propositional knowledge. But if we turn to the related distinctions brought up in the article, nonconceptual content does not seem to be a clear-cut case of implicit knowledge either. It cannot fulfill the requirement on verbal expressibility, because by definition it is not verbal. But what about accessibility and being under voluntary control?

Let us consider some cases in which the knowledge that lies behind the behaviour seems to rely on nonconceptual content. Examples of nonconceptual content as used in guiding behaviour while one's attention is attracted to something else (e.g., riding a bike) fit the description of procedural knowledge, which is governed by a rule that can only be active or inactive and is not open to scrutiny.

On the other hand, in the case of cycling, and perhaps even more in cases like playing tennis or golf or dancing, these activities can be deliberately improved. Different techniques can be tested, details can be changed, and the repertoire extended.

What is more, the standards that govern these activities are not only correctness conditions, that is, those that spell out whether representations match or fail to match their sources or targets, but also normative rules, or norms, which concern the quality of what is done. The same goes, for example, for craftmanship.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

The norms can be intersubjective, although it is impossible to formulate them explicitly. We can judge quantitatively measured properties according to explicit criteria. Other properties that influence our judgement of performances or products are experiential and not readily verbalisable, but they are nevertheless intersubjectively recognised. Examples of this can be found in judgements or classifications made in sports like gymnastics or figure skating, and also in judgements of style.

Craftsmanship as such does not depend primarily on the kind of conceptual, context-independent, and general knowledge representations D&P use in their theory. Instead, the activity is tuned to the context. It relies on a constant perceptual evaluation of the process (where perceptual is taken to involve all the senses) and not on verbal reflection. During this progressive evaluation, the subject incessantly makes decisions about what to do the next moment.

As an example, take a blacksmith or the architect working with clay models. The skill required to create new products constitutes a practical knowledge accessible to the subject and not reducible to procedural knowledge.

Moreover, not only people working with design or art make use of the external world in reasoning. Idiosyncratic representations tuned to what things are like or how they appear to the subject, rather than accurate conceptual descriptions, are prevalent. We pay attention to them, although we do not, and cannot, verbalise them. They often underlie decisions about the immediate future. But nonconceptual, contextual representations are not fit to enter into long-term planning or reasoning. They do not stretch into the distant future.

D&P's description of visually guided behaviour (e.g., the remarks about its being based on feature-placing instead of identification of objects) fits rather well with the activities governed by nonconceptual content described here, but a difference is that D&P seem to hold that visually guided behaviour is procedural and inaccessible. This is exactly what I would contest.

To sum up, I believe that the picture that D&P give of cognition is too crude. There are not only two opposing forms of knowledge, implicit and explicit, but also another kind that has properties from both sides, but also some properties of its own. The question is whether it can be incorporated into D&P's model.

Implicit representation, mental states, and mental processes

Richard A. Carlson

Department of Psychology, Penn State University, University Park, PA 16803-3106. **cvy@psu.edu** **gandalf.la.psu.edu/Rich/**

Abstract: Dienes & Perner's target article constitutes a significant advance in thinking about implicit knowledge. However, it largely neglects processing details and thus the time scale of mental states realizing propositional attitudes. Considering real-time processing raises questions about the possible brevity of implicit representation, the nature of processes that generate explicit knowledge, and the points of view from which knowledge may be represented. Understanding the propositional attitude analysis in terms of momentary mental states points the way toward answering these questions.

The theory outlined by Dienes & Perner (D&P) constitutes a significant advance in thinking about implicit knowledge. In particular, their analysis of knowledge states in terms of content and noncontent aspects, and the observation that there is at least a kernel of explicit content in even "maximally implicit" knowledge, provide important bases for sharpening the debates about implicit knowledge. Their contribution might be even clearer if its focus were inverted, addressing questions about what minimally explicit knowledge is sufficient to account for the observed performances of experimental participants in procedures presumed to reveal implicit processes.

Realizing the contribution of this theory, however, will require a kind of elaboration mostly neglected in the target article – a realization of the proposed distinctions in detailed processing terms, with a concomitant consideration of the time scale of propositional attitudes as they are realized in actual mental processes. As presented – and as typically discussed in the literature cited by D&P – a propositional attitude is a timeless entity, mostly considered apart from moment-to-moment mental activity. A similar set of distinctions can be made in an analysis that considers self, attitude (or "mode," Carlson 1997), and content as aspects of momentary mental states, embedded in processing streams comprising series of such states (Carlson 1997). Such mental states might be identified with individual goal states or the execution of single productions in computational theories like ACT-R, requiring perhaps several hundred milliseconds (Anderson & Lebiere 1998). Alternatively, hypotheses about their time scale might be based on a survey of perceptual and attentional phenomena (Pöppel 1988), suggesting that individual mental states have durations of about 30 msecs to 3 sec. The point is that if the set of implicit/explicit distinctions suggested by D&P is applied to brief, occurrent mental states, the implicit or explicit status of particular aspects of those states may also be brief. For example, a mental state whose explicit content includes "bachelor" is likely to be quickly followed by a state in which "unmarried" is no longer implicit but is part of an explicit content.

An important question that could be addressed by a focus on fine-grained processing details concerns how explicit knowledge is generated. Consider predicate-implicit knowledge, said to be sufficient to account for performance in subliminal perception tasks (sect. 2.1.1), or in implicit memory tasks (sect. 4.2). Theoretical parsimony suggests that the processes that generate explicit representations of (only) properties in these cases are also involved in generating fully explicit (or at least more explicit) knowledge in corresponding "explicit knowledge" situations. And presumably, further (later, in the case of perception) processes are also involved. But what might the nature of these *explicitness processes* be? D&P give us only a few hints, suggesting (a) that in the cases of development or relatively long-term learning, explicitness involves reflection, and (b) that in the case of perception some processes "downstream" (sect. 5) may produce more explicit representations.

One possibility is that reflection can be analyzed as an inferential process that takes as premises minimally explicit representations and relatively permanent (or permanently-available, for example, supported by always-active perceptual processes) explicit knowledge of one's own perceptual and cognitive processes. Or, currently instantiated goals together with minimally explicit representations might serve as premises for inferring more explicit representations (cf. Searle's 1983 analysis of the causal self-reference of intentional action; also see Carlson 1997). Of course, such accounts must be elaborated to explain why more explicit representations only sometimes result in situations that would seem to allow them. Alternatively, in some cases explicit representation (e.g., of attitude) might be supported by further processing that has the formal structure of hypothesis testing based on conditional reasoning – for example, if a representation really resulted from perception, then perceptual resampling should produce a representation with matching content. When it fails to do so (which is the case in the subliminal perception situation), an observer cannot infer explicit knowledge of the attitude "perceive." In other cases, processes such as associative priming – as in the "bachelor" example mentioned above – may quickly make additional content explicit. Finally, one might suppose that the structure suggested by the analysis of propositional attitudes serves as a kind of schema, with slots that have different criteria (e.g., of activation) for instantiation. Note that in any of these cases, the time scale of mental events is crucial – the explicit or implicit status of any aspect of knowledge may be attributed only to particularly, possibly brief mental states involved in controlling the performance to be explained. The broader attribution of implicit knowledge, as in

typical analyses of artificial grammar learning, must therefore constitute a claim that none of these explication processes has occurred, and perhaps that they cannot occur.

A process-based consideration of D&P's distinctions such as the one suggested here could help clarify some problematic aspects of their proposal. In particular, a puzzling near-contradiction in their article might be resolved. It seems odd to say that something "is represented purely implicitly," (sect. 4.4.2) because on their theory whatever remains implicit is *not represented at all for the participant,* though it will be represented *in a theory* that attributes implicit knowledge to that participant. Implicit representation may do important work for the theorist, but does no concrete work at all for the participant, reflecting instead the attribution of a disposition to use explicit knowledge in particular ways. As D&P's analysis of the relation between the implicit/explicit and unconscious/conscious distinctions suggests, it is a source of great confusion to think of implicit knowledge as a kind of representation that is explicit but unconscious or not accessible. Unpacking their distinctions in moment-to-moment process terms will help clarify the distinction between theorists' and participants' points of view, points of view with different implications for the status of particular hypothesized representations. Finally, understanding these points of view might help us understand in process terms what it means for a self to be related in a particular way to a proposition.

Explicitness and predication: A risky linkage

Andrew Carstairs-McCarthy

Department of Linguistics, University of Canterbury, Christchurch, New Zealand. **a.c-mcc@ling.canterbury.ac.nz**
www.ling.canterbury.ac.nz/adc-m.html

Abstract: Dienes & Perner (D&P) link explicit knowledge of facts to predication. But predication is basically a linguistic notion. Their approach therefore makes it difficult to attribute knowledge of facts to nonlanguage-users, such as animals. The explicit/implicit distinction, as D&P formulate it, is accordingly of little use for exploring the cognitive capacities of nonhuman primates – despite the increasing evidence for sophisticated social awareness among apes, implying mental representations of events in which participants are clearly distinguished. A revised formulation, less biased toward syntax as it happens to have evolved in humans, could avoid this drawback.

Dienes & Perner (D&P) hope that their distinction between explicit and implicit knowledge will be useful in a variety of research areas. One pertinent area is the comparative study of cognition in different species. For that purpose, we will need the criteria for explicit knowledge to be independent of language. An explicit/implicit distinction will be useless there if it restricts explicit knowledge to language-users by definition; it will be of only limited use if it is biased toward language-users in such a way that nonhumans, just by virtue of not being language-users, satisfy the criteria for explicit knowledge only with difficulty. D&P's formulation is not guilty of the first defect, but seems guilty of the second. It may be possible to reformulate it, however, so as to reduce this prohuman bias.

In section 2.1.1, D&P characterize knowledge, both explicit and implicit, as a propositional attitude. The content of a propositional attitude, in turn, involves an individual, a property, and a predication associating the property with the individual. Now, "predicate" is primarily a grammatical notion: its traditional counterpart is "subject," and a noun phrase and a verb phrase, fulfilling "subject" and "predicate" functions, respectively, are traditionally the essential components of a simple, nonelliptical, declarative sentence. So, because nonhumans do not use grammar, it may seem that D&P's formulation entails that nonhumans are incapable of propositional attitudes and so, a fortiori, are incapable of knowledge, even of an implicit kind. And surely, one may think, such a conclusion shows that there must be something wrong with the arguments or definitions that led to it.

D&P's view does not lead us quite to that conclusion, but close enough to warrant unease. According to their hierarchy of explicitness (sect. 2.1.1), the third element in the content of a propositional attitude, namely, predication, may be left implicit whereas only the first element (the property) or the first and second (the property and the individual) are explicit, but not vice versa. Even if predication crucially introduces grammatical structure to which only humans have access, there is scope for animals to have explicit knowledge of properties and individuals, at least. This still leaves important kinds of explicit knowledge inaccessible to animals. According to D&P (sect. 3.3), declarative knowledge (e.g., of the fact that this is a cat) presupposes explicit knowledge not just of a property (e.g., catness) and an individual (e.g., this animal), but also of predication (e.g., "this animal is a cat" as the content of a propositional attitude) and of factuality (the correspondence between this proposition and a real state of affairs); yet explicit knowledge of factuality, being higher than predication on the explicitness hierarchy, is denied to any creature that is incapable of explicit predication. This should make us uneasy because of the ample evidence that nonhuman primates have mental representations of their world, and in particular, of their conspecifics' behavior, in terms of predicate-argument structure – that is, they know who is who and who has done what to whom (Cheney & Seyfarth 1990; Dunbar 1996; Goodall 1986). Some nonhuman primates are apparently even capable of deliberate deception, which presupposes the ability to discriminate between facts and nonfacts (Byrne & Whiten 1988; Whiten & Byrne 1997). These discoveries are hard to reconcile with the weight that D&P place on predication.

In response, one might argue that other primates can indeed predicate explicitly, even without overt linguistic expression, in that their mental representations are expressed in a "language of thought." This argument might work if one could be confident that a language of thought shared with other species must exhibit the kind of syntax found in spoken human languages, distinguishing two kinds of complex expression, "noun phrases" and "sentences," with "predicates" as essential components of sentences but not noun phrases, and with sentences as the only candidates for truth or falsity. But there are no grounds for such confidence. A semantic predicate and its arguments can be expressed syntactically in a noun phrase, such as (2), just as readily as in a sentence, such as (1):

(1) Caesar defeated Pompey.
(2) Caesar's defeat of Pompey.

Why, then, does actual human syntax make available these two kinds of expression? A superficially attractive reason, namely, that only sentences can express facts, was cogently dismissed by Frege (1980); the truth expressed by (1) could just as well be expressed by (2), in a version of English where free-standing noun phrases are conventionally understood to refer to facts. Such a version of English even exists in the usage of those newspapers whose subeditors permit headlines such as *Cancer research breakthrough,* which are syntactically noun phrases rather than sentences. So the fact that the syntax of all or nearly all languages distinguishes between sentences and noun phrases is a genuine puzzle. A solution that invokes the distinction between "truth" and "reference" risks presupposing precisely the grammatical distinction that it purports to explain. A radically different solution (Carstairs-McCarthy 1998; 1999) appeals to factors specific to the evolution of humans, namely, bipedalism and the lowering of the larynx. I will not try to justify that proposal here, but merely point out what it implies for the distinction between implicit and explicit knowledge.

Dienes & Perner's distinction between knowledge of an individual (step 2 in their hierarchy of explicitness) and knowledge of the truth of a proposition (step 4) is parochial, appropriate to

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

knowledge as expressed in human language but not to knowledge in general. So the topmost point in the explicitness hierarchy need no longer be regarded as inaccessible in principle to nonhumans. What nonhumans can actually achieve cognitively is of course an empirical issue. But in recognizing the pro-human bias in "predication" at step 3 in the hierarchy, we can remove a possible source of pro-human bias in the study of comparative cognition.

Is factuality a matter of content?

Gregory Currie

Department of Philosophy, Flinders University, Adelaide, South Australia 5001. **greg.currie@flinders.edu.au**
coombs.anu.edu.au/Depts/RSSS/Philosophy/PhilosophyHome.html

Abstract: Dienes & Perner argue that there is a hierarchy of forms of implicit knowledge. One level of their hierarchy involves factuality, where it may be merely implicit that the state of affairs is supposed to be a real one rather than something imagined or fictional. I argue that the factual or fictional status of a thought or utterance cannot be a matter of concept, implicit or explicit.

I can utter the sentence "Martians have landed" as part of an account I am giving you of what happened today, or as part of my performance in a sci-fi movie. In the first case I make an assertion; in the second case I engage in fiction-making. Exactly what that second thing amounts to is controversial, but it seems highly likely that the difference between assertion and fiction-making is in some way a difference of emphasis; the difference is not in what is said – inasmuch as the same thing was said both times – but in the force with which it is said.

This is one way to make the factual/fictional distinction. Dienes & Perner (D&P) may agree, but they do want at least to add something to this account of the distinction. What they add plays an important role in their account of implicit knowledge and is, I believe, wrong. They claim that the factual/fictional distinction can be expressed as a difference of content. That might seem to be undermined by the example above, where two utterances differ with respect to the distinction but have the same content. D&P will reply that although the utterances have the same explicit content, they differ with respect to implicit content. According to their view, if I say "this is a cat," I have failed to make explicit where this is a fact in the real world, or just something about "a cat in some fictional context" (sect. 2.1.1, para. 2). If my statement is factual, its factuality belongs to the implicit content of what is said.

A point of clarification first. If factual contrasts with fictional, as D&P intend, then the factual is not the same as the true: Something can be false without being fiction. But D&P sometimes write as if the factual were identical to the true (e.g., sect. 2.1.1, para. 10). I will ignore this and assume that the factual is the asserted, or the seriously meant, or the intended-to-be-taken-as-true, or some such. That way we have a genuine contrast with the fictional.

Is the factuality of a thought or utterance a matter of implicit content? No. I can say, as part of a fictional performance, not merely, "this is a cat," but, "it really is a fact about the real world now that this is a cat," and still my utterance is fictional. Fictional utterances are generally about the real world, but they are about how someone imagines or asks us to imagine that it is, rather than about how it actually is. A fiction can be about President Clinton even though not everything it says is true of Clinton, and a fiction can be about the real world without everything in it being true of the real world. You can say anything in fiction that you can say outside it, and your saying it will not undermine the fictional status of the utterance.

D&P might respond by saying that this shows only that the implicit content of an utterance can be wrong. There might be no King of France, yet, in their account, "the King of France is bald" implicitly asserts that there is one. And "there is a cat" may implicitly say that there really is a cat, even if it turns out that this utterance is fictional. But there is really no parallel here. Asserting that the King of France is bald misfires (as people used to say) when there is no King of France. But authors who decide that emphasis is needed and write the line "there is a cat, and that's a fact about the real world" for one of their fictional characters have produced no error beyond a certain neglect of idiom.

This shows that the factual/fictional distinction is deeper than the distinctions of force between ordinary speech acts, say, question and request. I can say "can you pass the salt?" and by way of clarification, "I mean that as a question and not as a request for the salt." That makes clear that it was, exactly, a question and not a request. But if I say "there is a cat" and add "and that's a fact about the real world," no comparable clarification is achieved; the whole thing may just be part of my fictional performance.

But can I not clarify a fictional utterance of "there is a cat" by adding "and by the way don't take me seriously, this is just a play I am acting in," or something like that? I can, but my clarification will not serve to make content explicit that was implicit in the original utterance. If I say "there is a cat" as part of a fictional performance, anything implicit in my utterance must have the same fictional status as what was said explicitly, otherwise it would not count as the implicit content of that utterance. But the rider "don't take me seriously, this is just a play I am acting in" has to be taken as an assertion if it is to do any clarifying. And it cannot be implicit in what I originally said, namely, "there is a cat," because what I originally said was not asserted.

I am not, of course, saying that the factual/fictional distinction is ineffable. We can think and talk perfectly well about the distinction, both in factual mode and in fiction. What we cannot do is capture the factual or fictional status of an utterance within the content, implicit or explicit, of the utterance itself.

Explicit representations in hypothetical thinking

Jonathan St. B. T. Evans[a] and David E. Over[b]

[a]*Centre for Thinking and Language, Department of Psychology, University of Plymouth, Plymouth, PL4 8AA, United Kingdom;* [b]*School of Social and International Studies, University of Sunderland, Sunderland SRI, United Kingdom.* **j.evans@plym.ac.uk** **david.over@sunderland.ac.uk**

Abstract: Dienes' & Perner's proposals are discussed in relation to the distinction between explicit and implicit systems of thinking. Evans and Over (1996) propose that explicit processing resources are required for hypothetical thinking, in which mental models of possible world states are constructed. Such thinking requires representations in which the individuals' propositional attitudes including relevant beliefs and goals are made fully explicit.

Our interest in Dienes & Perner's (D&P's) target article relates to the idea that performance on higher level reasoning and decision-making tasks reflects the operation of distinct implicit and explicit systems of thinking (Evans & Over 1996; see also Reber 1993; Sloman 1996; Stanovich 1999). Implicit thinking has been characterised as relatively fast, high in processing capacity, and connectionist-like, associated with pragmatic processes that serve to focus attention on relevant features of the problem and retrieve associated knowledge from memory. The influence of such processes is relatively impervious to experimental instructions and inaccessible to verbal report. Explicit thinking, on the other hand, is relatively slow and sequential, and severely limited in processing capacity, but it is responsive to instructions, observable in think-aloud protocols, and allows the operation of volitional strategies (see Evans, in press). There is evidence that performance on tasks requiring explicit thinking processes is correlated with measures of general intelligence, whereas the performance on those that can be solved pragmatically is not (Stanovich & West 1998). All

the above characteristics suggest that explicit thought processes pass through (verbal) working memory, whereas implicit ones do not.

Evans and Over (1996) argue that the explicit thinking system is unique to human beings and related to language and reflective consciousness, but without specifying how and why this relationship has come about. This is where D&P's ideas are particularly helpful. They argue that explicit cognitive processes are simply those that require certain forms of explicit representation of knowledge. However, it is not just facts that need to be represented, but the individual's attitude toward those facts, including relevant beliefs, desires, and goals. For example, in the section on voluntary control (sect. 3.4), they discuss supervisory attentional systems (SAS), which is an idea closely related to what we call "explicit thinking." D&P argue that voluntary control provides a need to distinguish explicitly between content and attitude: "the SAS must be (or contain) a second-order mental state (one that represents desires)." This suggests the reason why the uniquely human possession of explicit thinking and reflective consciousness is permitted by the uniquely human possession of language. Essentially, explicit representations of this kind are constructed through language.

Evans and Over (1996) spend some time considering the function of explicit thinking. Why have we evolved this slow, sequential form of thinking when most of our needs can be satisfied by rapid and powerful implicit processing? Our analysis suggests that conscious resources (precious working memory space) are required for what we call "hypothetical thinking." Hypothetical thinking requires representations of possible states of the world and is necessary for anticipating and preparing for novel states of affairs. Deductive reasoning is hypothetical when its premises are not actual beliefs, but rather assumptions or suppositions. We wish to give an account of this reasoning in terms of mental models, though we do not fully accept any current mental models theory (Evans & Over 1997; Johnson-Laird & Byrne 1991). The use of deduction and induction in forecasting requires hypothetical thinking because we must construct a representation of some possible future state of the world. Consequential decision making consists of forecasting a number of possible future world states and representing the possible actions available to respond to these states. Scientific thinking is itself hypothetical when entertaining hypotheses about the way the world might be and deducing their consequences for making predictions.

D&P's analysis helps us understand why the explicit thinking system is required for such hypothetical thinking. Implicit decision making essentially involves responding to the immediate state of the world on the basis of previous experience, which we take to be learning embodied in neural networks. What all the kinds of hypothetical thinking discussed above have in common, however, is the need to represent what is possible, rather than what is actual. This distinction cannot be captured through implicit representations. Although reasoning researchers have generally focussed on the content of mental models – that is, what they represent – the process is clearly critically dependent on what D&P describe as the propositional attitude toward that content. When engaged in hypothetical thinking it is not sufficient for us to represent the content of what we know or believe. What is represented in hypothetical thought has to be kept distinct from what is known or believed with any confidence, because hypothetical thought is directed toward what is possible and not what is taken to be actual. Hypothetical thought uses suppositions or assumptions, and these must be represented as such, lest their content be confused with knowledge or firm belief about what is actually the case.

Consider trying to cope with changes in the world we have never experienced by supposing that global warming will continue. From this we may be able to infer that sea levels will rise and be prepared with appropriate action in response. The close connection between hypothetical reasoning and language can be seen in the fact that there is a linguistic form that can be used to sum up this reasoning: the conditional. That is, at the end of our hypothetical reasoning about global warming, we believe and may assert a conditional conclusion: that sea levels will rise if this warming continues. Conditional beliefs and assertions, whether made by ourselves or authorities and experts we respect, are obviously of great use to us in responding to new states of affairs that may arise. However, we are not satisfied with any current psychological account of the conditional (Over & Evans 1997), and finding a better account will be necessary before hypothetical reasoning and explicit thought can be fully understood.

In conclusion, we welcome the analysis of explicit knowledge provided by Dienes & Perner. Our only significant concern with the target article is the lack of reference to any work on the psychology of thinking. We hope that we have been able to show in this brief commentary that the concepts they advance bear directly on the role of explicit cognitive processes in tasks requiring hypothetical thinking.

Conceptual multiplicity and structure

Norman R. Gall

Department of Philosophy, University of Winnipeg, Winnipeg, MB, Canada, R3B 2E9. **gall@uwinnipeg.ca** **www.uwinnipeg.co/~gall**

Abstract: Dienes & Perner make three mistakes in their account of the "natural language meaning" of implicit-explicit knowledge: They fail to take the multiplicity of use of a concept seriously enough, they arbitrarily separate use of a concept and its conceptual structure, and they tend to tailor their analysis for use by the Representational Theory of Knowledge.

Dienes & Perner (D&P) make the claim that their analysis of the implicit-explicit distinction (with respect to knowledge specifically) will proceed from the "ordinary language meaning" (sect. 1) of these terms. It is my contention that they fail to go far enough in their analysis and, had they done so, much of the rest of the target article would have to be reconsidered. My three comments will be restricted to this point.

The insight that what competent speakers of the language say about these terms is critical, but D&P move from saying that "if something is conveyed but not explicitly, then we say it has been conveyed implicitly" (sect. 1) to this something being analysed as a "propositional attitude" (sect. 2.1). But the standard use of the term "implicit" by the ordinary language speaker does not usually embrace the notion of a "propositional attitude" or any similar notion. When we talk about implicit messages or notions in standard speech, we tend to talk about what sorts of things the speaker assumes the hearer already knows about or what anyone should be expected to know about without having it spelled out for them. The question as to what knowledge consists in does not enter the picture. We judge a speaker to be "assuming too much" if what is taken to be implicit is excessive or if what is spoken would be more appropriate for a more learned audience. Competent speakers are not concerned with a philosophical analysis of what knowledge consists in; they care about what should be taken as knowledge at any one time and whether they are saying straight out what needs to be understood or allowing the context to carry some of the burden. A more extensive exploration of what competent language users take as the meanings of "implicit" and "explicit" is required – a better overview of the concepts deployed here in terms of their actual use.

D&P correctly point out that sources of implicitness include both the use of the term in the language and the conceptual structure of the term itself, but they fail to see the connection between these two sources. The conceptual structure, those other allied concepts that must be understood alongside the term in question, cannot be separated from the use of the term in the language. The meaning, understanding, and explanation for all of these terms is logically inseparable and D&P seem to be saying something like:

"the use and the structure of the concept are somewhat independent; one could understand the conceptual structure of a term without making reference to the way the term is deployed by competent language users." But the error here seems to be a misunderstanding of how "contextual fashion/use" (sect. 1) cashes out. It seems correct and straightforward that the person who claims the King of France is bald also implicitly claims there is a King of France. However, D&P say that "in this sense, he did (and thus we say: "implicitly") convey that there is a King of France" (sect. 1). This may seem a small point, but it is not clear to me that the speaker intended to "convey" that there is a King of France. It may have been assumed that the hearer knew this and that it was not necessary to convey such trivial (though false) information. The important point is that there is more than one way to characterise what is going on in D&P's example and by ignoring this they fail to recognise the multiplicity of uses to which a concept can be put.

D&P conclude that in both cases (and if I am right above, this is no surprise) implicitly conveyed information "concerns *supporting facts* that are *necessary* for the explicit part to have the meaning it has" (sect. 1). However, they then go on to claim that "in our analysis the distinction is between which parts of the knowledge are explicitly represented and which parts are implicit in either the functional role or the conceptual structure of the explicit representations" (sect. 1). This move belies a shift from being concerned with the uses to which these utterances are being put by the speaker to a kind of reification of utterances as representations. There was no question of this sort of characterisation in the rest of the discussion. It appears that these necessary supporting facts need to be somewhere for D&P and this seems to me to be a mistake.

Nevertheless, this exposes something about the way D&P are analysing these concepts. It seems to me that they are already importing the Representational Theory of Knowledge (RTK) and applying it to their analysis of the natural language use of the terms. I would think that we would want to keep RTK far away from our analysis of these terms until we begin our account of RTK to avoid limiting what we can say about expressions of explicit and implicit knowledge. By slanting the analysis in its final stages, by not recognising the internal relationship between the conceptual structure of a term and the large number of different uses to which it is applied by language users, and by failing to analyse what these uses actually are, we are going to be unable to hand over to the RTK a sound understanding of these concepts. If we are unsure before applying these concepts in such a theory, any derived hypotheses will be conceptually questionable.

How does implicit and explicit knowledge fit in the consciousness of action?

Nicolas Georgieff[a] and Yves Rossetti[b]

[a]*Institut des Sciences Cognitives, 69500 Bron, France;* [b]*Espace et Action, INSERM Unit 94, 69500 Bron, France.* **rossetti@lyon151.inserm.fr**
www.lyon151.inserm.fr/unites/094_rossetti.html

Abstract: Dienes & Perner's (D&P's) target articles proposes an analysis of explicit knowledge based on a progressive transformation of implicit into explicit products, applying this gradient to different aspects of knowledge that can be represented. The goal is to integrate a philosophical concept of knowledge with relevant psychophysical and neuropsychological data. D&P seem to fill an impressive portion of the gap between these two areas. We focus on two examples where a full synthesis of theoretical and empirical data seems difficult to establish and would require further refinement of the model: action representation and the closely related consciousness of action, which is in turn related to self-consciousness.

Our first point concerns the sensory side of consciousness. Dienes & Perner (D&P) propose that for a given knowledge access to explicit representation depends mainly on making the elements of this knowledge explicit. The question accordingly arises whether all implicit information can, at least in principle, become explicit, particularly in the content of the dorsal-ventral dissociation (Goodale & Milner 1992) frequently referred to by D&P. Action offers a way to express implicit knowledge without declarative processes; this contrasts with the views of Millikan (1984) and Dretske (1988) (cited at the beginning of sect. 3.3) without being a form of procedural knowledge (as suggested at the end of sect. 3.3).

D&P note (sect. 5) that "sometimes the unconscious and conscious representations will contradict each other." What is the status of these instances of contradiction with respect to the accessibility of implicit knowledge to consciousness? D&P's reply is based on work by Clements and Perner (Perner & Clements 1997; 1999), who report a similarity with the dissociation observed in the visual system, but with respect to the theory of mind rather than to knowledge about objects used for simple actions. If we stick to the level of simple actions, the basic difference between the two visual pathways is clearly not based on consciousness. Not only do several types of implicit processing take place in the ventral stream, but it is the nature of the signals elaborated within each of these visual networks that seems to be the most relevant. The two visual systems elaborate specific representations that can be distinguished by their anatomical substrates (dorsal vs. ventral), their aim (action vs. identification; Milner & Goodale 1995), their spatial reference frame (egocentric vs. allocentric; Bridgeman 1992; Milner & Goodale 1995), their content (partial vs. global processing; Jeannerod & Rossetti 1993; Milner & Goodale 1995) and their life-time (short-lived vs. sustained; Rossetti 1998), as well as their access to consciousness.

Just as D&P describe how knowledge used by implicit learning can become explicit (end of sect. 5), so one can ask how the simple knowledge elaborated in one of the visual systems can access what is in the other. Several attempts have been made to stimulate information transfer between the action system and the identification system in healthy subjects, as well as in brain-damaged patients, such as those with blindsight and "numbsense" (reviewed in Rossetti 1998). The content of the declarative representation of the object seems to be transferable to the motor system, but this process has nothing to do with making implicit knowledge explicit. All experiments suggested that simple rendering of the motor representation explicit is not feasible; properties of the motor representation that can be expressed directly by a simple action toward an object cannot be expressed by other responses, such as verbal report and delayed action. This is certainly consistent with the fact that some of the motor representations are short-lived. Although it is mentioned by D&P, this point calls for more detail, because similar differences could also be found between implicit and explicit knowledge for mental processes outside the domain of action, thereby providing instances of "inaccessibility" of implicit knowledge (sect. 3.3). In this case D&P's model would have to make room for knowledge that cannot be made explicit.

Our second concern is with the action side of consciousness. D&P analyse several neuropsychological dissociations between explicit and implicit processes. However, other pathologies show a more specific dissociation between representation of the self (metarepresentation) and consciousness. Cognitive changes observed in neuropsychology (such as productive symptoms in anosognosia) and psychiatry (such as the delusion of alien control in schizophrenia) are very informative here. Schizophrenia can be considered a specific pathology of agency offering a striking illustration of a dissociation between different aspects of consciousness of action, in which a self-produced action can be correctly perceived and described but at the same time, systematically misattributed. The so-called positive symptoms displayed by schizophrenic patients are essentially disorders of self-consciousness and action consciousness or agency. These symptoms include thought insertion, auditory-verbal hallucinations, delusions of reference, and delusions of alien control. These false beliefs lead to a feeling of depersonalization, impairing the distinction between the self and the external world (Schneider 1959). In auditory hal-

lucinations, the patient hears voices that are experienced as coming from an external entity (Chadwick & Birchwood 1994). With other positive symptoms, too (e.g., thought intrusion, delusions of alien control, paranoid delusions, and reference delusions patients may declare that they are being acted on by alien forces, as if their thoughts or actions were controlled by external agents (Frith 1992).

Positive schizophrenic symptoms offer a specific model of dissociation between different aspects of explicit representation or knowledge of self-generated actions. They suggest a dissociation between the explicit representation of the content (or consciousness) of action, and the explicit representation of agency (i.e., reflective representation or metarepresentation). According to D&P, metarepresentation corresponds to explicit representation of the self as the holder of an attitude and it relies on an explicit representation of the attitude, that is, higher-order thoughts. We suggest that metarepresentation should also be considered as a form of consciousness of action. It is by generating metarepresentation that the self can become aware of its own actions. This is how one's mental productions (thoughts or representations; i.e., inner reality) are distinguished from the perception of external events (external reality) and how one's actions are distinguished from those of other people (i.e., how the self is distinguished from other selves). These properties relate to how an action is attributed to its proper origin, in other words, how a subject can make a conscious judgement about who is the agent of that action (an agency judgement).

D&P make a broader use of the "what" and "how" systems in vision. At a higher level, subjects may accurately attribute the origin of an action to themselves, yet ignore many aspects of their motor performance (Fourneret & Jeannerod 1998). This suggests that there are dissociable levels in actions with regard to access to consciousness. The signals used for controlling motor execution would be different from those used for generating conscious judgements about an action. Questions accordingly arise about the possible cognitive systems underlying the explicit or metarepresentational levels of knowledge of action.

We suggest that in schizophrenic symptoms the dissociation between the explicit content of action and its metarepresentation does not correspond to the classical dissociation between implicit procedural and explicit declarative knowledge (such as in blindsight), and could not be considered as simply included in the classical distinction between the explicit declarative system (ventral path) and the implicit visuo-motor system (dorsal path). Recent work has shown that the dorsal-ventral dissociation can be tracked further rostrally, up to the prefrontal cortex. Dorsal and ventral inputs to this structure seem to be segregated (Rossetti 1998). In addition, according to Frith (1992; 1995), disconnection between prefrontal (action command) and posterior associative areas result in a failure to anticipate sensory reafferents resulting from action, which may then be misattributed (Frith 1992; 1995). This suggests that more complex brain networks seem to be involved in consciousness of action than in conscious perception. These data deal directly with the agency problem.

Analysis of brain activity during several forms of action (active, passive, mentally simulated) has revealed a network common to all these conditions, to which the inferior parietal lobule (area 40), part of the supplementary motor area (SAM), and the ventral premotor area contribute (Jeannerod 1994; 1997). This cortical region is somewhat homologous with monkey ventral area 6 where one can record from "mirror neurons," not only when the animal performs a specific goal-directed hand movement (e.g., a grasping movement), but also when the immobile monkey watches the same movement performed by a conspecific (Rizzolatti et al. 1996a). Rizzolatti has proposed that monkeys recognize a motor action by matching it with a similar action motorically coded in the same neuronal population. We have suggested that this system could be a framework for studying dysfunctions of the mechanisms for answering the question "Who?" (e.g., the schizophrenic alteration of agency; Georgieff & Jeannerod 1998). This mechanism could be for our relationships with other individuals the counterpart of the "What?" and "Where?" mechanism for our relations with objects. To summarize, the reflective representations allowing the self to adopt a holder attitude require a representation of others. The implications of such a social system need to be developed in Dienes & Perner's model.

Does the hand reflect implicit knowledge? Yes and no

Susan Goldin-Meadow[a] and Martha Wagner Alibali[b]

[a]*Department of Psychology, University of Chicago, Chicago, IL 60637;* [b]*Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213.* **sgsg@ccp.uchicago.edu alibali@andrew.cmu.edu**
www.ccp.uchicago.edu/faculty.shtml
www.psy.cmu.edu/psy/faculty/malibali.html

Abstract: Gesture does not have a fixed position in the Dienes & Perner framework. Its status depends on the way knowledge is expressed. Knowledge reflected in gesture can be fully implicit (neither factuality nor predication is explicit) if the goal is simply to move a pointing hand to a target. Knowledge reflected in gesture can be explicit (both factuality and predication are explicit) if the goal is to indicate an object. However, gesture is not restricted to these two extreme positions. When gestures are unconscious accompaniments to speech and represent information that is distinct from speech, the knowledge they convey is factuality-implicit but predication-explicit.

Dienes & Perner (D&P) make an excellent case that the distinction between implicit and explicit knowledge is many-layered. The challenge comes in finding the most useful way to characterize the layers. We focus here on the distinctions made within the "Content" box (Fig. 1), using them to identify the layer that best characterizes knowledge expressed in gesture.

Spontaneous gestures at times convey information that differs from that conveyed in the speech they accompany (Church & Goldin-Meadow 1986; Goldin-Meadow 1997; Goldin-Meadow et al. 1993). In section 4.3., D&P suggest that when gesture conveys information different from the information conveyed in speech, it reflects "thoughts about reality that have not yet been recognized as being about reality" – in short, gesture is factuality-implicit.

D&P draw a parallel between gesture-speech discordance and the dissociation between the knowledge bases underlying children's understanding of false belief. At a time when children display no understanding of false belief in their verbal responses they demonstrate a reliable understanding of false belief in their visual orienting responses (Clements & Perner 1994). Children *look* at the place where the protagonist thinks an object has been moved, even though they fail to *say* that this is the correct place. D&P argue that such visual orienting responses are factuality-implicit.

Two problems arise. First, this analysis puts gesture on a par with visual orienting responses. On intuitive grounds, this seems incorrect because gesture is symbolic, eye glances are not. Second, in Clements and Perner (1994), the pattern of gesture was not like visual orienting responses but like speech. When asked where the protagonist would look, children indicated the incorrect place with *both* words and gestures.

D&P's framework can be used to resolve both problems. We begin by considering gesture in relation to visual orienting responses. We agree that both may be factuality-implicit (although we return to this question below). We suggest, however, that gesture differs from eye glances at the level of prediction – gesture may be predication-explicit, whereas eye glances are not. Information that is "useable by different parts of the system" (sect. 4.3) is predication-explicit. We offer two examples to show that spontaneous gestures can meet this criterion.

First, when asked to describe algebra word problems that they have read, adults sometimes convey different information in gestures and speech. In such cases, adults subsequently solve the

problem using a strategy compatible with their spoken description 32% of the time. But 43% of the time, they solve the problem using a strategy compatible with their *gestured* description (Alibali et al. 1999). In these instances, the information expressed uniquely in gesture "previews" the subsequent problem solution. Thus, gesture represents information that can be referenced by different parts of the system.

Second, children often express strategies for solving mathematical equations in gesture that they do not express in speech (Alibali & Goldin-Meadow 1993; Perry et al. 1988). When later asked to rate the correctness of solutions generated by different problem-solving strategies, children rate solutions generated by strategies that they conveyed uniquely in gesture *higher* than solutions generated by strategies they did not express at all (Garber et al. 1998). Thus, the information children convey uniquely in gesture is not tied to the hands but can be accessed by other systems – gesture is consequently predication-explicit.

We are currently attempting to ask the crucial question – is information conveyed uniquely through eye glances also accessible to other systems? Is it predication-explicit? Using an eye-tracker, we are able to determine the pattern of eye glances children produce when asked to solve equations. If children convey patterns through visual orienting behaviors that are *not* found in either their speech or gesture, we can then ask whether the patterns unique to eye-glances predict their subsequent ratings as well as patterns that are unique to gesture (cf. Garber et al. 1998). We suspect that they will *not* – that visual orienting behaviors will not be predication-explicit, and thus will be distinct from spontaneous gesture.

We now return to factuality. We agree with D&P that spontaneous gesture is factuality-implicit – that is, speakers do not recognize their gestured comments as being about reality. One way to test this claim is to encourage speakers to be aware of their gestures. When gestures are truly spontaneous, they sometimes tap knowledge that cannot be expressed in words. If speakers are made aware of their gestures, this could change – gesture should become factuality-explicit, and should no longer convey different information from speech.[1] Indeed, in Clements and Perner (1994), children were asked to indicate where the protagonist would look, and many responded by pointing. These children were aware of having gestured. Gesture and speech were therefore both factuality-explicit (as well as predication-explicit) and, perhaps as a result, patterned together.

To summarize, gesture does not have a fixed position within D&P's framework. Instead, its position depends on the nature of the knowledge it expresses. If the goal is to move a pointing hand to a target (a visually guided movement[2]; Bridgeman 1991; Bridgeman et al. 1997, sect. 4.1), neither factuality nor predication is explicit, and the knowledge reflected in gesture is fully implicit. In contrast, if the goal is to indicate an object (a declarative act; Clements & Perner 1994), both factuality and predication are explicit, and the knowledge reflected in gesture is therefore explicit. However, gesture is not restricted to these two extreme layers of the D&P framework. When gestures are unconscious accompaniments to speech and represent information that is distinct from speech, the knowledge they convey is factuality-implicit but predication-explicit.

In some contexts, spontaneous gestures access a knowledge base that is distinct from the knowledge base that informs speech. Gesture may be abstracted from perception or action (e.g., Alibali et al. 1998) but is not itself perception or action. Hence, gesture extends beyond knowledge that is embedded in action. However, gesture is not recognized as being about reality, and is therefore not fully explicit. We argue that gesture reflects an important waystation in the progression from implicit to explicit knowledge – one that offers unique insight into implicit thought.

ACKNOWLEDGMENTS

Work described in this commentary was supported by Grant RO1-HD18617 from NICHD and by a grant from the Spencer Foundation to S. Goldin-Meadow. We thank M. Morin for helpful comments.

NOTES

1. We thank John Cacioppo for suggesting this study.

2. It is important to point out that visually-guided behaviors and visual orienting behaviors do not always pattern in the same way (although they appear to do so in false belief tasks, cf. Clements & Perner 1996). A salient example is the young infant's knowledge of objects, which appears more sophisticated when measured via visual orienting responses (Baillargeon 1987; Spelke et al. 1992) than when measured via visually-guided reaching (Piaget 1954). Bertenthal (1996) suggests that this discrepancy can be resolved by acknowledging two distinct knowledge bases – one that subserves the perceptual control and guidance of actions, and one that subserves the perception and recognition of objects and events. As far as we can tell, the D&P framework does not capture this distinction – both knowledge bases are fully implicit. It might be worth incorporating into the framework a dimension that could distinguish the two.

Implicit knowledge in engineering judgment and scientific reasoning

Michael E. Gorman

Department of Technology, Culture, Communications and Systems Engineering, University of Virginia, Charlottesville, VA 22903.
meg3c@virginia.edu **repo-nt.tecc.virginia.edu**

Abstract: Dienes & Perner's theoretical framework should be applicable to two related areas: technological innovation and the psychology of scientific reasoning. For the former, this commentary focuses on the example of nuclear weapon design, and on the decision to launch the space shuttle Challenger. For the latter, this commentary focuses on Klayman and Ha's positive test heuristic and the invention of the telephone.

Dienes & Perner (D&P) outline four areas of application of their ideas to research. In this brief commentary, I want to add a fifth: the psychology of science (Feist & Gorman 1998) and technology (Gorman 1998). Consider an example. Mackenzie and Spiniardi (1995) argue that a great deal of implicit knowledge is involved in the development and maintenance of nuclear weapons and that much of this knowledge may be lost when the current generation of weapons designers retires. Similarly, Gusterson has written about the "group of senior scientists who have experienced many nuclear tests and who therefore "really understand" how the weapons work. Other scientists speak of these men as irreplaceable, because so much of their knowledge is tacit knowledge that is not, and probably cannot be, written down" (Gusterson 1996, p. 106).

D&P remind us of the degrees and types of explicitness a statement like, "this is a nuclear weapon" might have. Scientists might know that this is an effective weapon now, and might know how they knew that – by what test, or facts, or evidence, or experience; this level of explicitness corresponds to all the levels in Figure 1 of the target article. Or scientists might know that this is not an effective weapon anymore, even if it was one in the past, and be unable to articulate precisely why they feel that way. In this case, the content box in D&P's Figure 1 would be explicit, but not the attitude box, because attitude includes justification. Or does attitude just incorporate that "feeling of knowing"?

Consider a different example. Roger Boisjoly had seen the damage to the O rings on the space shuttle from previous flights, and was sure that they would blow at the low temperature projected for the Challenger launch. But the data he presented were ambiguous (Vaughan 1996). Boisjoly's judgment was hard to make explicit – when pressed, all he could say was that the decision to launch was a "step away from goodness." Boisjoly appeared to be at level 1 in Table 1 – he knew that the shuttle was not safe at this particular temperature, but could not articulate all the reasons behind his judgment.

Like many experts, Boisjoly and the weapons designers have to make judgments under uncertainty. Boisjoly did not know that the Challenger would blow up if launched; he merely felt that this kind of a disaster was significantly more likely at a lower launch

temperature (Vaughan 1996). The D&P model does not seem to accommodate this kind of uncertainty.

Boisjoly's implicit knowledge and that of the weapons designers could be considered procedural, in the sense that both are concerned with applications. Should the shuttle be launched at a low temperature? Should a nuclear warhead be subjected to further tests? D&P seem to distance themselves from Singley and Anderson's (1989) view that procedural knowledge is encoded declaratively; theirs is closer to a position like Bechtel and Abrahamsen's (1991). Boisjoly's intuitions about the O rings that caused the Challenger's failure were based, in part, on the look and feel of these rings after previous launches (Vaughan 1996), and hence could not have been previously encoded as a set of formal, declarative rules.

Confirmation and disconfirmation. D&P's framework might be applicable to resolving an issue in the literature on confirmation and disconfirmation in scientific reasoning. Klayman and Ha (1987) argue that much of this literature confounds confirmation with what they call a positive test heuristic. Consider an example. Alexander Graham Bell had a hypothesis that what he called an undulating current was the only possible way of transmitting speech. He contrasted it with an intermittent current, the one commonly used in telegraphy, which he said could not transmit speech. Following Klayman and Ha, a positive test would involve Bell's building an apparatus that created the undulating current. If his hypothesis was right, then the device should transmit speech. If it did not, then this positive test would result in a disconfirmation.

But Bell knew that he was not a skilled electrician. Therefore, even when he designed a positive test and obtained a negative result, he did not see it as a falsification of his hypothesis; instead, he kept on trying to achieve the positive results. Faraday did this in his early experiments on magnetic induction – he knew the magnet he was using was not very powerful, so he was happy to get an occasional positive result (Tweney 1985).

D&P might argue that in an explicitly confirmatory test, the intention to confirm, the expectation of obtaining a positive or negative result, and the experimenter's confidence in the reliability apparatus and methods would be explicit. If any of these elements were implicit, one could not be sure the test was explicitly confirmatory. Bell kept detailed notebook entries concerning his experiments, but he did not discuss his intentions at this level of detail. I have therefore argued that Bell followed an implicit confirmation heuristic (Gorman 1995).

If Bell were alive today, we could "protocol" him to find out whether he had explicit intentions that he did not write down (Ericsson & Simon 1984). Would Dienes & Perner agree that using a combination of notebooks and protocol analysis is a reliable way of determining the extent to which heuristics and judgments are implicit? These authors should consider how their framework might be extended to group problem-solving situations (Dunbar 1995; 1997), where implicit, shared mental models play an important role (Levine & Moreland, in press).

The functional role of representations cannot explain basic implicit memory phenomena

Yonatan Goshen-Gottstein

Department of Psychology, Tel-Aviv University, Ramat Aviv, Israel, 69978.
goshen@freud.tau.ac.il

Abstract: The propositional account of explicit and implicit knowledge interprets cognitive differences between direct and indirect test performance as emerging from the elements in different hierarchical levels of the propositional representation that have been made explicit. The hierarchical nature of explicitness is challenged, however, on the basis of neuropsychological dissociations between direct and indirect tests of memory, as well as the stochastic independence that has been observed between these two types of tests. Furthermore, format specificity on indirect test of memory challenges the basic notion of a propositional theory of implicit and explicit knowledge.

Dienes & Perner (D&P) present a propositional theory for understanding the distinction between implicit and explicit knowledge so as to provide a common ground for the use of this distinction in different fields of investigation. One of the most interesting arguments the authors make is that there are hierarchical constraints on explicitness (see sect. 2.1.1; Table 1; Fig. 1), so that if certain elements are represented explicitly (e.g., a particular individual), elements that are lower in the hierarchy must also be represented explicitly (e.g., property). If elements that are lower in the hierarchy are explicitly represented, however, then higher elements may or may not be explicitly represented.

This general argument allows D&P to interpret cognitive differences between direct and indirect test performance as emerging from elements, in different hierarchical levels of the propositional representation, that have been made explicit. For example, in section 2.1.1, D&P suggest that subliminal presentation of target materials results in explicit representation of the property of a stimulus (low level), but not of the fact that there is a particular stimulus event that is of that kind (higher level). That subjects cannot verbally report the occurrence of the subliminal stimulus is explained by the need, on direct tests, for a (high-level) explicit representation that there is a particular stimulus event. The sensitivity of indirect tests to the subliminal stimulus is explained by the influence of the property of the representation (low-level fact) that was made explicit. Comparable arguments are made regarding memory (sects. 3.2; 4.2), with successful performance on indirect tests of memory only requiring representation of low-level elements (e.g., property), but direct tests requiring that high-level elements (e.g., factuality) also be explicitly represented (sect. 3.2).

Unfortunately, a propositional account of implicit and explicit knowledge is insufficient to explain implicit memory performance because of the pattern of neuropsychological dissociations, experimental dissociations, and stochastic independence that is observed between direct and indirect tests of memory. Because of the empirical findings that will now be described, the unique quality performance on direct and indirect tests of memory has to be determined, and the different underlying memory systems that mediate performance on the two types of tests must be uncovered (e.g., Schacter & Tulving 1994).

The key empirical findings that the propositional theory should have trouble interpreting are now described. First, it is unclear how the neuropsychological double dissociation between the memory systems underlying direct and indirect tests could be explained by the theory. Gabrieli et al. (1995) contrast performance by two types of patients. Amnesic patients, with damage to medial temporal lobe structures, do not display memory on direct tests, yet show normal performance on indirect tests of memory. According to D&P's theory, these patients are unable to make explicit high-level elements of representations (such as factuality). This idea can somehow be fathomed. The reverse pattern, however, seems to falsify the theory. A patient with a lesion to the right occipital lobe was found to exhibit unimpaired performance on a direct test. Thus, high-level elements of representations (e.g., fac-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

tuality and temporal context) were explicit in this patient. From the hierarchical nature of the theory, if high-level elements are explicit, low-level elements must be explicit too, and the patient should be unimpaired on indirect tests. Hence, the impaired performance of this patient on indirect tests presents a real puzzle for the theory.

Equally troubling are the predictions that the theory would make with regard to the ability of successful performance on a direct test to predict performance on an indirect test. If the two tests were administered successively, then successful memory of an item on the direct test would suggest that the high-level elements of the representation for that item were made explicit. Because indirect tests can benefit from elements that have either been made explicit or not, the probability that the item will be produced on the indirect test, conditional on its having been remembered on the direct test, is higher than had it not been remembered on the direct test. This prediction has been disconfirmed. Tulving et al. (1982; Hayman & Tulving 1989) found stochastic independence between word recognition (direct test) and a subsequent word-fragment completion test (indirect test).

Finally, according to D&P, what determines bona fide performance on an indirect test is implicit representation of the elements of a fact (or elements of the attitude or self) that constitute part of the proposition. Presumably, the propositional nature of the representation should be insensitive to format of presentation. Yet format of presentation seems to be a critical factor in predicting implicit memory performance. For example, on tests such as perceptual identification (e.g., Jacoby & Dallas 1981) or word-fragment completion (e.g., Tulving et al. 1982), where subjects are required to identify a visually degraded display, indirect memory performance is diminished, or completely eliminated, if the similarity of retrieval cues (e.g., word fragments) to studied items is reduced by crossing the modality of presentation between study and test (e.g., Jacoby & Dallas 1981; for a comprehensive review, see Roediger & McDermott 1993). Moreover, even when study and test presentations are within the same modality, presenting different study and test materials such as words and pictures (e.g., Weldon 1991) or words in different languages (e.g., Kirsner & Dunn 1989) has been shown to reduce performance on indirect tests. It is unclear how a propositional theory can account for these findings.

Implicit and explicit knowledge: One representational medium or many?

James A. Hampton

Department of Psychology, City University, London EC1V 0HB, England.
j.a.hampton@city.ac.uk **www.city.ac.uk**

Abstract: In Dienes & Perner's analysis, implicitly represented knowledge differs from explicitly represented knowledge only in the attribution of properties to specific events and to self-awareness of the knower. This commentary questions whether implicit knowledge should be thought of as being represented in the same conceptual vocabulary; rather, it may involve a quite different form of representation.

Implicit knowledge is characteristic of most human cognition (and, as far as one can tell, of *all* animal cognition). If a proper account could be given of levels of implicit representation, it would therefore have tremendous explanatory power and would open up a way to understanding numerous problems in cognitive science.

A proper distinction between explicit and implicit knowledge is important in the study of conceptual knowledge. When interrogated about the contents of their conceptual knowledge, it is well known that people generate variable and idiosyncratic responses (Rosch & Mervis 1975). For example, in one unpublished study, I examined the relation between the relative importance that people attach to criterial properties of a concept and their judgements of the relative typicality of instances of the same concept. Subjects performed two tasks. The first was to rank order a set of properties in terms of how relevant they were to the definition of a category. The second was to rank order a set of category instances in terms of their typicality. The data were analyzed to measure the similarity between individuals on either task. If people have explicit knowledge of the reasons why they consider some instances more typical than others, and if there is any individual variability amongst the population (as could reasonably be expected), then the similarity of a pair of individuals on one task should be related to their similarity on the other task.

When the two sets of similarities were compared, however, the pattern of similarity between individuals in terms of the centrality of attributes showed no correspondence at all with the pattern of similarity between individuals in terms of instance typicality. It appears then that much of our conceptual knowledge is implicit.

If conceptual understanding is implicit, then the critical question will be how the representational language of explicit knowledge is grounded in implicit knowledge. The challenge is to provide a semantics for knowledge representation with the flexibility of the different levels of explicit/implicit awareness. Is the conceptual representational language the same at different levels of the system, and is it only the predication of properties to objects or events and to the self as knower that differentiates the levels? This would appear to be D&P's view. Or should the representation of knowledge using a vocabulary based on natural language be restricted to explicit levels of representation?

Fodor (1998) has argued strongly against the grounding of explicit concept terms (such as bird or bachelor) in a more implicit set of semantic features or roles. To Fodor, the meaning of the word "bird" is just BIRD – a conceptual atom that is grounded through its symbolic relation to the class of birds in the real world. We may learn that certain propositions hold of birds in general (e.g., that birds are creatures), but this set of propositions – whether necessarily true or not – is not constitutive of the meaning of the concept.

In section 2.2., D&P suggest that an atomic, nondecomposable representation may be thought of as having an implicitly represented property structure. For example, whereas "bachelor" can be decomposed into its component features, on any particular occasion it may be used in an explicit representation without being decomposed. A person may be able to claim, "I knew that I was looking for a bachelor, but I had neglected the fact that the person would have to be unmarried." Yet there is clearly a major difference between this type of atomism and the type advocated by Fodor. Fodor's arguments for conceptual atomism suggest that there is no implicit property structure encoded at some deeper or more hidden level – there is just an informational semantic connection to the class of bachelors, and the possibility of learned generalisations that one could make about the class.

The problem becomes more apparent if one asks that information one would wish to include in the implicit conceptual structure of a representation, and how this information might be constrained or determined. D&P suggest that implicit conceptual structure involves "necessary supporting facts." The closest they come to giving a detailed account of these is when they state, "Using 'bachelor,' oneself commits one quite strongly to 'male' and 'unmarried' lest one show oneself ignorant of the meaning of the word bachelor in the language spoken" (sect. 1).

But in using the term, one is also committed to an indefinite number of other propositions such as "not a vegetable" or "composed of cells containing DNA," while in addition one is committed (to a greater or lesser extent) to all the more prototypical aspects of being a bachelor, such as living alone, wariness of marriage, or fondness for solitude. There is no simple logical way of selecting those aspects of a concept's meaning that should be considered as forming the implicit conceptual structure, from, on the one hand, the indefinitely large number of necessary inferences that follow from the concept, and, on the other hand, the many probabilistically related attributes that characterise so much of our conceptual knowledge.

A clear account of the implicit/explicit distinction with respect to conceptual content is needed in cognitive science. In making connections across disparate fields of cognition, Dienes & Perner have drawn attention to the possibility of offering a unifying account of the distinction, which would have far-reaching consequences. It remains to be seen, however, whether it makes sense to think of the implicit representation of knowledge making use of the same language-like representational medium as is found in explicit conceptualisation.

Making implicit explicit: The role of learning

Bruce D. Homer and Jason T. Ramsay

Centre for Applied Cognitive Science, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, M5S IV6, Canada.
{bhomer; jramsay}@oise.utoronto.ca
www.oise.utoronto.ca~bhomer/Bruce_homer.html

Abstract: Three forms of implicit knowledge are presented (functional, structural, and procedural). These forms differ in the way they are made explicit and hence in how they are represented by the individual. We suggest that the framework presented by Dienes & Perner does not account for these differences.

Dienes & Perner (D&P) present a framework for conceptualizing the nature of mental representations that attempts to capture the various natural language uses of *implicit* and *explicit* knowledge. Although D&P find several points of agreement between the different uses of *implicit*, we suggest that they do not adequately capture the nature of "fully implicit" knowledge; hence essential, qualitative differences inherent in the different uses of *implicit* are lost in the D&P framework. There are at least three forms of implicit knowledge: structural, functional, and procedural. The differences between them become apparent when one considers what is needed to make them explicit.

One form of implicit knowledge derives from "property-structure" implicitness (sect. 2.2) in which an explicitly represented property (e.g., "bachelor") is a compound of two or more basic properties (e.g., "unmarried" and "male"). Property-structure implicit knowledge is semantically related to explicit knowledge: One cannot use the word "bachelor" correctly without knowing that it means "unmarried" and "male." For knowledge that is structurally implicit to become implicit, an individual need only consciously reflect on the implications of the explicit knowledge. The explicit property (e.g., "bachelor") acts as a heuristic for recalling the implicit properties and so on individual can easily provide the longer version of the heuristic (i.e., "unmarried and male"). A heuristic represents implicit knowledge in a way that makes it the most available to conscious or explicit representation.

Contextual function is another source of implicit knowledge. As an example of this, D&P point out that certain propositions (e.g., "the present king of France is bald,") presuppose other propositions (e.g., that there is a present king of France). The presupposition is therefore implicit in the first proposition. Presuppositions are given as the "prime case" of contextual function implicitness (sect. 1, para. 6). A similar source of implicit knowledge, not addressed by D&P, is *entailment*. Two or more propositions, when related according to a set of semantic rules, can entail certain other propositions. These entailed propositions are implicitly contained in the original propositions and the semantic rules. For example, in Plato's *Meno* (1986), through the process of questioning a slave boy about geometry, Socrates succeeds in eliciting the Pythagorean theorem. This is a sense of "implicit" that is not easily accounted for in D&P's framework. The logical propositions and the rules by which they are related to create the theorem are explicitly known to the slave boy; it is the way in which they are explicated that is new. This is a unique instance of implicitness. In the case of "bachelor," the implicit constituents are made explicit through the efforts of the individual. In the case of entailments, they are not, although they are recognized as being logical explications. Once explicated, entailments are immediately grasped by the individual, although their previous existence was not explicitly represented.

A final example of implicit knowledge is procedural knowledge. Certain information (e.g., a rule, theory, or concept) is contained implicitly in any procedure. For example, children who are able to balance odd-shaped blocks on a beam have a naïve theory of torque implicit in their balancing procedure (Karmiloff-Smith 1992). For this implicit information to become explicit, however, simply telling an individual the implicit information is not enough. Specific concepts may have to be learned so that children will reflect on their procedure and explicate their theory. In a series of studies, Piaget (1976) investigated children's explicit representations of their actions and found that there is a lag between their ability to perform actions and their ability to describe how they perform these actions. For example, children demonstrated great skill in performing tasks that require centrifugal force (e.g., hitting a target with a slingshot); however, the children's representations of how they succeeded on the task and the actual means by which they achieved the result were discrepant. This is a case where implicit knowledge (the procedure) and explicit representations conflict. How do children become aware of this implicit knowledge? Piaget suggests that this is through the process of "reflexive abstraction," which entails developing new conceptual structures that allow the emergence of this reflexivity. Furthermore, the development of these new conceptual structures may depend on extrinsic factors. Homer and Olson (1999), for example, have found evidence that literacy is responsible for children becoming aware of certain linguistic properties of their speech.

In the examples above, we have presented three different forms of implicit knowledge. One of the essential ways in which these forms differ is in how the implicit knowledge can be made explicit. For structurally implicit knowledge, an individual need only engage in conscious reflection to explicate the implicit knowledge. For functionally implicit knowledge, an individual must be told the implicit knowledge (e.g., presupposition or entailment); however, once told, the implicit knowledge is immediately grasped (i.e., explicitly known) by the individual. For procedurally implicit knowledge, individuals must learn new concepts that can be used to reflect on their procedure. The key point is that these different forms of implicit knowledge become explicit in very different ways. This suggests that they are represented by the individual in qualitatively different ways. We suggest that any framework attempting to capture the nature of knowledge representations must account for these differences.

Fishing with the wrong nets: How the implicit slips through the Representational Theory of Mind

Luis Jiménez[a] and Axel Cleeremans[b]

[a]*Department of Psychology, Universidad de Santiago, 15706 Santiago, Spain;* [b]*Cognitive Science Research Unit, Université Libre de Bruxelles CP 122, 1050 Brussels, Belgium.* **jimenez@usc.es** **axcleer@ulb.ac.be**
web.usc.es/~psljim/englishpage.html
srsc.ulb.ac.be/axcWWW/axc.html

Abstract: Dienes & Perner's target article is not a satisfactory theory of implicit knowledge because in endorsing the representational theory of knowledge, the authors also inadvertently accept that only explicit knowledge can be causally efficacious, and hence that implicit knowledge is an inert category. This conflation between causal efficacy, knowledge, and explicitness is made clear through the authors' strategy, which consists of attributing any observable effect to the existence of representations that are as minimally explicit as needed to account for behavior. In contrast, we believe that causally efficacious and fully implicit knowledge exists, and is best embodied in frameworks that depart radically from classical assumptions.

The goal pursued by Dienes & Perner (D&P) in this target article is an ambitious one, as they aim to build a theory of implicit and explicit knowledge that would enable cognitive scientists to distinguish among the diversity of senses in which one can consider knowledge to be "implicitly" held by a cognitive system. This kind of conceptual effort is commendable in and of itself, given that common-sense and technical terms are often blended in maturing disciplines such as cognitive science. D&P's proposal is therefore to be welcomed, even if, as we will show through our commentary, their effort tends to be conceptually misguided. To put it simply, we believe that D&P have dropped a conceptual net that is ill-suited for bringing back the intended fish.

Consider D&P's strategy – their "conceptual net." This consists of analyzing knowledge as a "propositional attitude according to the representational theory of mind" (sect. 1), a theory that assumes a representation "constitutes knowledge if it is put in . . . a *knowledge box* or . . . data base" (sect. 2.1). D&P propose a functional distinction between explicit and implicit knowledge, according to which the content of any knowledge (i.e., the content of the representations in the knowledge box) is explicit if it is a function of that representation to indicate precisely that content. In contrast, information is implicit if it is conveyed only as an indirect consequence, or supporting fact, of what is explicitly represented. The authors offer this functional criterion as a way of independently distinguishing between different types of knowledge that shape human cognition. However, as the rest of the target article clearly illustrates, such a functional criterion is not independent of the criterion of knowledge efficacy, and hence results in a disturbing conflation between causal efficacy, knowledge, and explicitness.

This conflation seems unavoidable in D&P's conceptual approach, given that the only external way to ascertain the function of some knowledge is to ascertain its cognitive and behavioral effects. In relying on this criterion the authors have no choice but to consider all the knowledge that produces observable effects as explicit knowledge *at the specific level that is minimally needed to account for the observed effects.* This logic is widely illustrated throughout the target article, and goes roughly like this: If you observe an effect that appears to depend on the presence of some knowledge, label this knowledge as explicit at that particular level, and let any other knowledge that appears to have no bearing on the observed behavior be implicit. If the efficacious knowledge includes properties of the stimulus, then call it "property" explicit; if it includes, for example, the holder of that knowledge, then call it "self" explicit.

Clearly, the problem with this strategy is that labeling as implicit only the knowledge that has no bearing on a specific situation, makes it somewhat odd even to consider it as "knowledge," no matter how tightly related to explicit knowledge this implicit knowledge might appear to be from an external point of view. Thus, to paraphrase a sentence that D&P apply to the criterion of accessibility (sect. 3.3), we might say that if implicit knowledge were not causally efficacious, then it would not qualify as knowledge and, in any case, there would be no evidence that there was any implicit knowledge at all. Implicit knowledge, then, in this framework, is the name for a nonentity – a fish that has slipped through D&P's conceptual net.

How did the fish slip through the net? We surmise that the problem lies in D&P's adoption of the "representational theory of mind." According to this framework, representations constitute knowledge if they appear in a "knowledge box." It is important to note that whether or not a particular representation enters the knowledge box is determined by whether "the representation is used as a reflection of the state of the world and not, as it would be, for example, if it were in a *goal* box, as a typically nonexisting but desirable state of the world" (sect. 2.1). Now, the only way for an outside observer to ascertain whether or not a particular representation that an agent has is in its knowledge box (i.e., is "used as a reflection of the state of the world") is to examine whether it influences the agent's performance on some task. But this reasoning is exactly identical to that entailed by the criterion used by D&P to ascertain whether knowledge was explicit at any given level. In other words, we seem to be caught in a maze of twisty little conceptual corridors that all point to the same conclusions: In D&P's framework, (1) a representation can only constitute the agent's knowledge if it is in the agent's knowledge box, (2) a representation can only influence performance if it is in the agent's knowledge box, and (3) any representation that is in the knowledge box is necessarily explicit in at least the specific way needed to account for observable behavior. It should be clear that this conceptual net has shark-sized holes in it.

Where do the holes come from? The problem, we surmise, is with the knowledge box. The notion that one can account for the way in which our representational systems are organized by assuming that representations are formed and put in databases of different kinds is simply inadequate to capture the dynamics of cognition. What is the alternative? We would eliminate the "knowledge box" as a requirement for the definition of knowledge and instead assume that representations can simultaneously constitute knowledge and be causally efficacious without ever being tokened in any way. For example, observing that "butter" has been perceived in a subliminal perception experiment because it exerts detectable effects on performance does not imply that the property of "butter" has been somehow unconsciously represented in the subject's knowledge box (as D&P strangely suggest in sect. 5) or, worse, that it has been represented in some unconscious zombie-like twin of the knowledge box. It simply means that the relevant neural pathways were activated sufficiently to bias further processing in the relevant direction when the stem completion or lexical decision task is actually performed. The knowledge embedded in such pathways is knowledge that is simultaneously causally efficacious and fully implicit. It does not produce any kind of conscious or unconscious "attitude" and hence cannot be accounted for by a representational theory of mind. Clearly, such knowledge is better captured through dynamical approaches such as the connectionist framework (see Cleeremans 1997; Cleeremans & Jiménez, submitted; [see also van Gelder: " The Dynamical Hypothesis in Cognitive Science" *BBS* 21(5) 1998]; Mathis & Mozer 1996; O'Brien & Opie, 1999) – a perspective with which D&P otherwise agree. A particularly important and difficult issue in this context is to chart the divide between processes and representations, but this is a matter for further debate.

ACKNOWLEDGMENTS

Axel Cleeremans is a Research Associate of the National Fund for Scientific Research (Belgium). This work was also supported by Grant XUGA21106B98 from the Xunta de Galicia to Luis Jiménez.

Memorial states of awareness versus volitional control: The role of task differences

Sachiko Kinoshita

Department of Psychology, Macquarie University, Sydney, NSW, Australia 2109. **sachiko.kinoshita@mq.edu.au**

Abstract: Dienes & Perner's analysis provides a clear theoretical justification for using a demonstration of volitional control as a criterion for conscious awareness. However, in memory tasks, the converse does not hold: A phenomenological awareness of a memory episode can arise involuntarily, even when the task does not require retrieval of the episode. The varying amounts of volitional retrieval required by different memory tasks need to be recognized.

Dienes & Perner (D&P) do much to integrate the distinction between implicit and explicit knowledge across different research domains. My commentary will focus on an issue that has been much debated in the domain of implicit memory, namely, the distinction between retrieval volition and memorial states of awareness. In section 4.2, D&P provide an analysis of the consequences of different elements being represented implicitly following a study episode. In my view, their analysis is somewhat mistaken, because the various consequences are all described in the context of a recognition test that requires retrieval to be intentional (i.e., retrieval volition is necessarily voluntary). As a result, the first two cases in which retrieval volition is involuntary are not captured well by the scenarios described.

I suggest instead that the first two cases in Table 2 are better illustrated with alternative examples. Let me focus on the distinction between the second and third scenarios, both involving recognition based on a feeling of familiarity. In everyday experience, an example involving involuntary volition (Scenario 2) would be a chance meeting with an old school friend that results in a feeling of familiarity that arises spontaneously; in contrast, an example involving direct volition (Scenario 3) would be when one is asked to identify the perpetrator in a lineup of suspects. The difference between the two lies in the fact that in the former, the memorial awareness arises spontaneously, even though it is not required by the task, whereas in the latter, it is demanded by the task.

In memory research, the status of a feeling of familiarity in an implicit memory test is hotly debated. Researchers working with normal subjects would be familiar with the case of having administered an ostensibly implicit test of memory (e.g., asking subjects to report the first word that comes to mind that starts with "but-" when one of the studied words was "butter"), and being told by subjects at a post-experimental debriefing that they were aware that the word they produced was in the study list.

Currently, two contrasting interpretations of this scenario can be identified. On the one hand, there are those who argue that to the extent that the memorial state of awareness (a feeling of familiarity) arose spontaneously, the scenario should be described as a case of involuntary conscious memory (e.g., Richardson-Klavehn et al. 1994; Schacter et al. 1989). Such researchers have also argued that to the extent that retrieval was involuntary, the priming effect observed in such a case should still be considered as representative of implicit memory.

In contrast, Jacoby and his colleagues (Jacoby 1991; Jacoby et al. 1993) have defined memory in terms of a dichotomy: One form of memory is aware and involves intentional control; the other form is unaware and automatic. Because this framework equates memorial states of awareness with retrieval volition, it defines involuntary conscious memory out of existence. Within the latter framework, then, the implicit stem-completion task in the above scenario is described as being "contaminated" by explicit memory.

D&P's interpretation of the relationship between volitional control and memorial states of awareness (sect. 3.4) generally aligns itself with Jacoby's framework. Their analysis within the representational theory of knowledge indicates that control of actions requires explicit representation of the content (of action), factuality (or lack thereof), and attitude. D&P accordingly argue that this justifies that "voluntary control is used as a criterion for consciousness" (sect. 3.4). However, in my opinion, this analysis makes the same mistake as Jacoby. Although voluntary control may be used as a (conservative) criterion for consciousness, D&P have not provided the justification for the converse. As illustrated by the example of a chance encounter with an old school friend, an explicit sense of pastness does not necessarily mean that retrieval was intentional. This memorial state of awareness may arise without deliberate retrieval (it is beyond the scope of this commentary to describe the mechanism for an involuntary feeling of familiarity – but see Moscovitch 1995). The two concepts should accordingly be kept distinct.

In conclusion, D&P provide an insightful discussion of various notions that have been brought into contact with the implicit-explicit distinction, including volitional control and consciousness. The representational theory of knowledge provides a sound theoretical basis for relating the two, something that has not been achieved previously. Something that is lacking in D&P analysis, however, is a recognition of the varying degrees of external support involved in (and hence conversely, the varying amounts of self-initiated cognitive operations required by) different tasks. For example, free recall provides minimal clues at retrieval (e.g., "Tell me the words you saw") and hence requires the retrieval cues to be generated internally. In contrast, an implicit word fragment completion task (e.g., "Complete this fragment: "-ys-e-y") contains perceptual cues that could guide processing.

Clearly, a greater amount of volitional control is demanded by tasks that depend on self-initiated processing. This is of course the fundamental assumption of the transfer-appropriate processing (TAP) framework (Kolers & Roediger 1984; Roediger 1990; Roediger & Blaxton 1987). Researchers working in the area of implicit memory, even if they do not believe that the TAP framework is sufficient to explain the difference between implicit and explicit memory, nevertheless acknowledge the need to take into account the differences between test tasks. I also believe that a complete account of the distinction between implicit and explicit memory requires the appreciation of differences between tasks and the amount of volitional control demanded by them.

Implicit and explicit learning in a hybrid architecture of cognition

Christian Lebiere[a] and Dieter Wallach[b]

[a]*Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213;* [b]*Institut für Psychologie, Universität Basel, 4056 Basel, Switzerland.*
cl+@cmu.edu act.psy.cmu.edu/ACT/people/lebiere.html
wawllachd@ubaclu.unibas.ch
www.unibas.ch/psycho/Rognitiv/wallach.html

Abstract: We present a theoretical account of implicit and explicit learning in terms of ACT-R, an integrated architecture of human cognition as a computational supplement to Dienes & Perner's conceptual analysis of knowledge. Explicit learning is explained in ACT-R by the acquisition of new symbolic knowledge, whereas implicit learning amounts to statistically adjusting subsymbolic quantities associated with that knowledge. We discuss the common foundation of a set of models that are able to explain data gathered in several signature paradigms of implicit learning.

In their target article, Dienes & Perner (D&P) present a conceptual analysis of knowledge with the goal of systematically integrating the diverse uses of implicit-explicit distinction in several research areas. The authors (1) provide a precise definition of the terminology, (2) link it to related distinctions, and (3) derive new empirically testable predictions from their theoretical framework.

D&P explore the implicit-explicit distinction in terms of the degree of explicitness of predicates. This symbolic knowledge representation account is linked in section 4.3 (Development) to a

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

connectionist framework to furnish an explanation of the statistical emergence of that knowledge. Whereas D&P's representational hierarchy is fundamentally a theory of the degrees of explicit knowledge, because neither the representation in terms of weights nor the learning mechanisms in terms of statistical rules provide any measure of explicitness. We argue in this commentary that only a hybrid architecture provides the necessary representational properties for a truly integrated account of explicit and implicit knowledge.

The ACT-R cognitive architecture (Anderson & Lebiere 1998) provides such a hybrid framework for a comprehensive theory of explicit and implicit knowledge. Although one is tempted to equate explicit knowledge with declarative knowledge (represented as chunks) and implicit knowledge with procedural knowledge (represented as production rules) in the ACT-R production system, we agree with D&P's contention in section 3.3 (Procedural versus declarative knowledge and accessibility) that, although procedural knowledge is not verbally accessible, the procedural-declarative distinction is not identical to the implicit-explicit distinction. Instead, we propose that ACT-R's other duality, between symbolic and subsymbolic knowledge, provides a convincing account of explicit and implicit knowledge.

With the exception of some connectionist approaches, research on implicit learning and knowledge has suffered from a lack of cognitive modeling that would have helped clarify the implicit-explicit distinction in terms of precise computational structures and mechanisms. We have developed detailed ACT-R models of the main paradigms relevant to that distinction, including sequence learning (Lebiere & Wallach 1998; in preparation), control tasks such as Broadbent's transportation task and sugar factory (Lebiere et al. 1998; Wallach & Lebiere 1998) and implicit memory (Anderson et al. 1998).

Focusing on declarative chunks that are associated through subsymbolic activation processes, this foundational claim of our models not only provides a unifying account of the implicit-explicit distinction across different fields of investigation, but also supports existing theoretical positions in the respective paradigms. In each model, explicit knowledge is acquired in the form of symbolic chunks containing learning instances: pairs of stimuli in sequence learning, a single input-output pattern in control tasks, and the association between a word and its letters in implicit memory. Because ACT-R is an activation-based architecture, the ability to retrieve these declarative chunks is controlled by their associated real-valued parameters. The ACT-R sequence learning model (Lebiere & Wallach 1998; in preparation) presents an example of this approach. In sequence learning, a particularly abundant research area in implicit learning, subjects are asked to react as quickly as possible with a discriminative response to stimuli presented sequentially in one of a number of locations on a computer screen. Unbeknownst to the subjects, the presentation of stimuli follows a systematic order. Learning of event sequences is accessed indirectly by comparing response latencies with systematic versus random sequences. Providing an integrated account, ACT-R has been able to explain a wide range of sequence learning experiments (e.g., Curran & Keele 1993; Perruchet & Amorium 1992; Willingham et al. 1989). Basically, the model encodes pairs of consecutive stimuli in declarative units, successively building up chunks that span larger fragments of the sequence. Whereas representations encoding sequence fragments form the model's explicit knowledge, subsymbolic associations between chunks are learned implicitly, allowing activation to spread from stimuli to sequence fragments and priming them from subsequent retrieval by production rules. Our approach is thus related to the suggestion of Sloman (1996) who views implicit learning as an associative process, whereas explicit learning is conceptualized as a rule-based mechanism. Generally, implicit and explicit knowledge are fundamentally linked in ACT-R: Chunks could not be retrieved from memory without sufficient support from underlying activation quantities; and those quantities cannot be directly accessed and would be meaningless without the associated chunks.

ACT-R models of implicit learning achieve many of the goals set out by D&P. For example, the implicit memory model provides a precise and detailed account of their priming data in section 4.2 (Memory). In that model, the context sensitivity required in section 4.4.1 (Predication) is achieved by identifying in each chunk the list to which the item belongs. The stochastic nature of judgment discussed in section 4.4.2 (Reflection on attitude) is present in each model because memory retrieval in ACT-R is a fundamentally stochastic process: Noise is added to each chunk's activation, determining their probability of retrieval.

In conclusion, we think that these models establish that the ACT-R architecture realizes the main objective stated by D&P in section 1 (Introduction: Objectives), which is to "create a common terminology for the use of the implicit-explicit distinction in different research areas." It provides a detailed computational account of the implicit learning characteristic of connectionist models together with the explicit representations detailed by D&P. As Newell long ago stated (Newell 1973), this is the proper role of integrated cognitive architectures. [See also BBS multiple book review of Newell's *Unified Theories of Cognition BBS* 15(3) 1992.]

What is the cat in complex settings?

Pierre-Jean Marescaux and Patrick Chambres

LAPSCO, U.F.R. de Psychologie, Université Blaise Pascal, 63000 Clermont-Ferrand, France. **{marescaux; chambres}@lapsco.univ-bpclermont.fr**

Abstract: Dienes & Perner present a hierarchical model that addresses the nature – implicit versus explicit – of knowledge in areas as diverse as learning, memory, and visual perception. This framework appears difficult to apply to complex situations, such as those involving implicit learning, because of the indeterminacy that remains regarding knowledge at the low-level in the hierarchy. These reservations should not detract from the positive features of this model. Among its other advantages, it is well adapted to priming phenomena in which the information responsible for the individual's behavior can be precisely defined.

According to Dienes & Perner (D&P), knowledge is represented by a partial hierarchy, in which the implicit-explicit distinction is made at each level. Their framework seeks to integrate, in a unified perspective, the often divergent uses of the implicit-explicit distinction, and to account for these concepts in research areas as diverse as learning, memory, and visual perception. In this commentary, we discuss the extent to which the framework is suitable for dealing with problems related to the area of implicit learning.

It should be noted that, in the area of implicit learning, the implicit-explicit distinction refers to either the process of learning, its results (knowledge), or both. A number of theorists have argued, however, that only the process of learning itself should be defined as implicit, not the knowledge that results from it, and that, consequently, the former should be the focus of attention, not the latter (e.g., Frensch 1998; Perruchet et al. 1997). Still, the paradox in such a position is that to label a process as implicit or explicit, it is almost always necessary to examine its output (Jiménez 1997). In this sense, the framework proposed by D&P is applicable to the area of implicit-explicit learning, even though it covers only part of it.

The issues that must be raised, then, concern the general applicability of the framework and its ability to promote significant progress. According to D&P, the boundary between the implicit and the explicit is determined by evaluating the level in the hierarchy at which knowledge is represented (i.e., content, attitude, self). This may be accomplished without difficulty in simple cases such as in demonstrations of priming where the researcher controls the nature of the information that is provided to the experimental participants (e.g., "doctor" or "butter"). In this case, it is possible, through either a bottom-up or top-down analysis, to identify precisely the highest level that can be labeled as explicit. On the other hand, D&P's approach may be more difficult to ap-

ply when the information to be learned is more complex (e.g., artificial grammar). To re-evoke an example used by D&P themselves, the critical question at issue is: What is the cat in complex settings?

Participants in implicit learning experiments must typically learn a complex set of stimuli, rules, or correlations that they cannot articulate (Berry & Dienes 1993). It must be admitted, moreover, that researchers typically do not know the representational nature of the knowledge that results from such learning. They must resort to a best guess about its nature. Within the field of artificial grammar learning, at least four accounts have been offered for the production of grammatical judgments. These rely on: (a) knowledge about the deep structure of the environment (Reber 1989), (b) similarity to stored exemplars (Vokey & Brooks 1992), (c) access to multiple pieces of fragmentary knowledge (Dulany et al. 1984), and (d) episodic-processing that offers a large potential number of responses (Whittlesea & Dorken 1993). However, research has not been conducted to choose decisively among these candidate frameworks. As a result, individual researchers use tasks compatible with the hypothesis that appears most plausible or that is actually under investigation. The tasks are typically ill-adapted to assess the knowledge that is actually available (see Shanks & St. John's information criterion, 1994). This recurrent problem is cause for concern as it is liable to impede the assessment of the explicit nature of knowledge.

It is indeed difficult to imagine how the implicit nature of knowledge at a given level of the hierarchy can be established if the knowledge at lower levels has not been clearly identified beforehand. To illustrate this, imagine a researcher who has observed that exposure to letter strings generated by an artificial grammar results in the classification of new grammatical and nongrammatical stimuli at significantly above-chance levels. This effect is now well known and unsurprising, but, at the very least, it provides support for existence of "property-implicit" knowledge. The problem is thus to ascertain whether such knowledge is explicit or implicit at the level of attitude. If the researcher assumes that the knowledge is abstract and that grammatical judgments are holistic, it will seem natural to answer the question using the grammaticality test. This time, however, one also asks participants to rate their confidence in each decision. Because no relationship between confidence and accuracy is observed, the researcher concludes that either the knowledge is attitude-implicit or that the participants do not know what they do, in fact, know.

Let us further assume that the judgments are indeed made in an analytical way and, based on elementary informational components (e.g., bigrams, trigrams, legal positions), each of them is attitude-explicit fragmentary knowledge. This assumption is partly supported by Perruchet and Pacteau's (1990) finding that bigram recognition confidence correlates with frequency of occurrence in the study phase (conducted with full strings). The procedure described earlier seems quite unsuitable for the problem because each confidence judgment produced on a full string must be computed, in a complex way no doubt, from elementary confidence concerning the perceived fragments. To come as close as possible to the available knowledge, the researcher should have used a procedure in which the confidence judgments concerned elementary informational components.

In sum, the framework proposed by D&P, when applied to implicit learning, fails to account fully for the processes of implicit learning. Furthermore, as applied to complex situations, the framework requires meeting a condition that is in fact quite difficult to meet, that is, knowing the exact nature of the knowledge that underlies the participant's behavior. Thus, in the domain of implicit learning, the framework is clearly limited.

Explicit knowledge in dolphins?

Eduardo Mercado III[a] and Scott O. Murray[b]

[a]*Center for Molecular and Behavioral Neuroscience, Rutgers University-Newark Campus, Newark, NJ 07102;* [b]*Institute of Theoretical Dynamics, University of California, Davis, CA 95616.* **mercado@pavlov.rutgers.edu**
www.cmbn.rutgers.edu/~mercado/welcome.html
smurray@itd.ucdavis.edu itd.ucdavis.edu/~murray

Abstract: The theoretical framework proposed by Dienes & Perner sets the wrong standards for knowledge to be considered explicit. Animals other than humans possess knowledge, too, some of which is probably explicit. We argue that a comparative approach to investigating knowledge is likely to be more fruitful than one based on linguistic constructs and unobservable phenomena.

We agree with Dienes & Perner (D&P) that there is no simple dichotomy between implicit and explicit knowledge and that the idea of characterizing knowledge along a scale of explicitness is worth considering in detail. The ambiguities associated with the explicit/implicit distinction and the need for more precisely defined classifications have been discussed extensively by other researchers (for a review, see Engelkamp & Wippich 1995). By decomposing knowledge in terms of parameters derived from the representational theory of mind, D&P hope to resolve these ambiguities and thereby to provide a precise theoretical system for describing levels of explicit knowledge that can be consistently applied across various fields. The usefulness of their framework is limited, however, in that: (1) the linguistic/anthropocentric constructs they use to define knowledge, as well as levels of explicitness, make their theory applicable only to adult humans, (2) the criteria they endorse for experimentally identifying various levels of explicit knowledge are overly dependent on linguistic competence and unobservable phenomena such as consciousness, volition, and intention, and (3) it is unclear how one might distinguish between the explicitness of representations and the explicitness of retrieval and reporting mechanisms using their framework.

Knowledge research has focused almost exclusively on adult humans. Even studies of knowledge in children have primarily been concerned with identifying when and how adult-like (i.e., verbalizable) knowledge develops. We suggest that a comparative approach to understanding knowledge can provide a broader perspective on how brains encode, maintain, and retrieve information. Within the field of comparative cognition, for example, knowledge is described in terms of learned relationships between neural events (Olton et al. 1992; Roitblat 1987). These events can be generated by external stimuli (e.g., by sights or sounds) or they can be internally generated (e.g., producing movements or memories). Not only does this framework allow knowledge to be analyzed independently of verbal reports but it can also be mapped onto specific brain subsystems (Eichenbaum 1997; Merzenich & deCharms 1996; Squire 1992). Studies of knowledge in nonhumans offer greater objectivity because they are less likely to be corrupted by either introspection or linguistically based intuitions.

Distinctions between explicit and implicit knowledge in nonhumans are seldom made in the scientific literature. This is probably because explicit knowledge is typically defined in terms of consciousness, and most researchers are hesitant to make the as-of-yet empirically unsupportable claim that any species other than *Homo sapiens* is "consciously aware." Instead, knowledge in nonhumans is often described as being either procedural or declarative (or as involving stimulus-response vs. stimulus-stimulus associations, or perceptual vs. conceptual representations, etc.). Animal cognition research focuses first on discovering what animals can represent (or know) and then on understanding how they represent this knowledge. D&P start with the assumption that knowledge is represented propositionally, and then seem to equate existence of a representation with explicitness of knowledge (e.g., in sect. 2.1.3, they suggest that if the self is represented propositionally as knowing, then knowledge of a fact is fully explicit). It seems unlikely that knowledge in nonhumans is exclu-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

sively propositional. Yet there is evidence showing that some nonhuman species use metaknowledge to guide their actions, suggesting that they can assess what they know.

For example, Smith et al. (1959) examined whether bottlenosed dolphins (*Tursiops truncatus*) would respond adaptively to uncertainty in a psychophysical test by escaping in the way that humans do. This study essentially asked, "Does the dolphin know when it does not know?" Smith et al. found that when a dolphin and humans were tested on an identical task their responses to uncertainty were nearly indistinguishable. These findings provide evidence that in certain situations dolphins may represent their own knowledge in ways comparable to humans. Do these findings provide evidence of attitude-explicit knowledge in dolphins? If not, why not? If so, does this imply that dolphins represent their uncertainty propositionally? Human subjects escaped when uncertain because that is what they were instructed to do. If asked why they escaped, they might say, "Because I was uncertain." Would this then show that the human subjects' knowledge was fully explicit? If alternatively, a subject said, "Because you told me to," does this imply that his knowledge is only content-explicit?

Similar issues arise in studies of deferred imitation and action repetition. Dolphins and humans have both shown the ability to reproduce actions (on command) that they have observed or performed in the past (Bauer & Johnson 1994; Mercado et al. 1998; 1999; Xitco 1988). How explicit is the knowledge involved in such tasks? Does deferred imitation involve (or require) conscious recollection, episodic memories, intentional retrieval of those memories, and voluntary reproduction of the recalled actions? Introspection might lead one to conclude that imitation of actions does involve such processes. Certainly if someone were to describe verbally the actions they had observed, this would be viewed as compelling evidence of fully explicit knowledge of the events. How would a reenactment be any less compelling? Yet when the organism doing the imitating is a dolphin, the tendency is to view these performances as merely showing evidence of innate abilities (i.e., fully implicit knowledge) rather than evidence of explicit knowledge. These examples illustrate some of the difficulties associated with determining what information is explicitly represented.

The explicit/implicit distinction was originally introduced as a way of distinguishing between different ways of retrieving or expressing knowledge (in particular, memories). D&P attempt to generalize this distinction to the representation or encoding of knowledge. Cognitive neuroscience studies have provided evidence that explicit retrieval tasks activate brain regions that are not activated in implicit tasks (Badgaiyan & Posner 1997). Differences in neural activation have also been observed in tasks involving deep versus shallow encoding (Buckner & Koutstaal 1998). Such neural correlates could greatly facilitate comparative research if nonverbal analogs of current implicit and explicit tasks could be developed. Theoretical descriptions of the relative explicitness of the encoding and retrieval of information will be much more useful when they can be grounded in empirical observations that are not species- or task-dependent.

Applying a theory of implicit and explicit knowledge to memory research

Neil W. Mulligan

Department of Psychology, Southern Methodist University, Dallas, TX 75275-0442. **mulligan@mail.smu.edu**

Abstract: This commentary discusses how Dienes & Perner's theory of implicit and explicit knowledge applies to memory research. As currently formulated, their theory does seem to account simultaneously for population dissociations and dissociations between conceptual and perceptual priming tasks. In addition, the specification of four distinct memorial states (correlated with different recognition test responses) faces important methodological challenges.

Dienes & Perner (D&P) present a compelling analysis of the relationship between implicit and explicit representations. This analysis has much to recommend it, beginning with the laudible goal of extracting commonality from disparate research areas, and continuing with a rich framework for investigating these commonalities. The application of this framework to research on memory requires additional clarification, however. Because the implicit-explicit memory distinction has been one of the most active research areas in cognition during the past 15 years (e.g., Roediger & McDermott 1993; Schacter 1987), this domain is of critical importance for a general theory of implicit and explicit knowledge. Unfortunately, with regard to implicit memory, the present theoretical treatment raises more questions than it answers.

A candidate theory of implicit and explicit memory must address the set of dissociations that have been obtained between measures of implicit and explicit memory, the majority of which have been obtained in research contrasting direct and indirect tests. Consistent with the notion that direct and indirect tests are capable of measuring different aspects of memory are numerous reports of population dissociations, in which participants from populations impaired on direct memory tests (such as people with amnesia, depression, or schizophrenia, and older adults) show normal or near-normal levels of repetition priming on indirect memory tests (e.g., Denny & Hunt 1992; Light 1991; Schwartz et al. 1993; Shimamura 1986; 1993). Converging evidence comes from experimental manipulations, such as the read/generate manipulation and study modality, which also dissociate performance on indirect and direct tests (see Roediger & McDermott 1993, for a review). A final relevant aspect of this literature is that dissociations also occur between indirect conceptual and perceptual tests. For example, dividing attention and levels-of-processing affects conceptual priming tasks such as category-exemplar production and free association but they typically have no measurable impact on perceptual tasks, such as word-fragment completion (e.g., Mulligan 1998; Roediger & McDermott 1993).

D&P apply their theory of implicit and explicit knowledge to memory in section 4.2. In light of the above discussion, how are we to integrate the hierarchy of representations in Table 2 with the literature on priming? The most straightforward way is to conceive of the hierarchy as a model of memory encoding varying from very impoverished, in which only the property is explicitly represented (e.g., the word form "butter" is represented explicitly) to richly encoded, in which all elements of the trace, including experiential source (i.e., origin) are explicitly represented. Such a conceptualization can accommodate population dissociations if one makes the reasonable assumption that amnesics (and to a much lesser degree, older adults, the depressed, and schizophrenics) have a deficit in creating explicit representations of origin and (presumably) factuality and time.

However, such a conceptualization has difficulty accounting for the relationship between conceptual and perceptual priming tasks. In particular, D&P associate the strongest form of implicitness in which only property is explicitly represented (i.e., the most impoverished form of encoding) with enhanced performance on conceptual priming tasks, proposing in section 4.2 that "the mere

presence of the word 'butter' can enhance the likelihood of answering a request to list dairy products with 'butter'," is a description of conceptual priming in the category-exemplar production task. A less impoverished encoding, with explicit representations of a feature compound (e.g., the word form plus perceptual details) is associated with performance on perceptual priming tasks such as perceptual identification. Given that greater elaboration during encoding produces less impoverished traces (i.e., more aspects of the trace are represented explicitly), this account implies (incorrectly) that manipulations of elaborative processing are less likely to affect conceptual than perceptual tasks. Specifically, the account implies that even the least elaborative encoding condition would likely give rise to the minimally explicit representation on which conceptual priming relies and that further opportunities for elaboration would not enhance this probability. In fact, just the opposite is found. Encoding conditions that limit the opportunity for elaborative encoding, such as divided-attention conditions and nonsemantic encoding tasks, reduce and often eliminate priming on conceptual tasks, but have little effect on perceptual priming (Mulligan 1998; Mulligan & Hartman 1996; Srinivas & Roediger 1990).

It seems unlikely, then, that the minimally explicit representation is associated with conceptual priming. The more natural assumption is the reverse, that conceptual-elaborative encoding leads to greater levels of explicit representation. This account also confronts a problem because inherent in this view is the notion that conceptual-elaborative processing is the core deficit in amnesia and other populations with compromised performance on direct tests. Contrary to this notion, amnesics (as well as older adults) typically show normal levels of performance on conceptual priming tasks such as category-exemplar production and free association (e.g., Light & Albertson 1989; Shimamura 1986; 1993), implying intact conceptual encoding processes. Parenthetically, it should be noted that intact conceptual priming in memory-impaired populations, coupled with the finding that conceptual priming tasks satisfy the retrieval-intentionality criterion (e.g., Mulligan 1996; Weldon & Coyote 1996), indicate that the conceptual priming tasks of category-exemplar production and free association are legitimate measures of implicit memory (i.e., they are not greatly influenced by explicit contamination).

One final point concerns the relationship between different levels of explicit representation and states of awareness, and the methodological challenge of measuring these states. D&P have expanded the distinction between the Know and Remember states of recognition (as have others, e.g., Conway et al. 1997) to include states leading to recognition by familiarity (based on compound-explicit and predication-explicit representations) and informed guessing (based on property-explicit representation). It remains a challenge to this theoretical framework to develop models for measuring these aspects of memory and demonstrating the necessity of positing four separate states of awareness correlated with qualitatively different underlying representations. There are reasons for pessimism. Consider the literature on the two-state, Remember-Know distinction. The distinction relies heavily on findings that R and K responses produce dissociations (see Gardiner et al. 1998; Rajaram 1996 for recent reviews). However, it has recently been demonstrated that a two-criterion signal detection theory (SDT), which posits a single underlying dimension of familiarity, can reproduce the dissociations between R and K responses taken as evidence for two qualitatively different memorial states (Donaldson 1996; Hirshman & Master 1997). In addition, a traditional criterion manipulation (instructions about the proportion of old items on the recognition test) affects both R and K responses in ways predicted by the two-criterion SDT model (Hirschman & Henzler 1998). Two points follow. First, R and K responses are affected by decision processes and therefore should not be thought of as unmediated reflections of underlying representations. Second, the two-criterion SDT model stands as a challenge to the Remember-Know method. One need not endorse a single-state familiarity model of recognition to argue that such a model (formalized in SDT) needs to be definitively ruled out before one accepts the estimates of Remember-Know from the more complex two-state model (cf. Gardiner et al. 1998). The present theory, proposing four states, faces an even higher standard of evidence.

Explicit factuality and comparative evidence

Shaun Nichols[a,b] and Claudia Uller[b]

[a]*Department of Philosophy, College of Charleston, Charleston, SC 29424;* [b]*Center for Cognitive Science, Rutgers University, Piscataway, NJ 08854.*
nichols@cofc.edu **uller@ruccs.rutgers.edu**

Abstract: We argue that Dienes & Perner's (D&P's) proposal needs to specify independent criteria when a subject explicitly represents factuality. This task is complicated by the fact that people typically "tacitly" believe that each of their beliefs is a fact. This problem does not arise for comparative evidence on monkeys, for they presumably lack the capacity to represent factuality explicitly. D&P suggest that explicit visual processing and declarative memory depend on explicit representations of factuality, whereas the analogous implicit processes do not require such representations. Many of the implicit/explicit findings are also found in monkeys, however, and D&P's account needs to explain this striking parallel.

According to Dienes & Perner's (D&P's) account, the distinction between implicit and explicit processes can be captured by appealing only to the content of the representations. This content-based account is an extremely bold and provocative alternative to the traditional view that explicit processes tap cognitive mechanisms that are not implicated in implicit processes. D&P make an important contribution to the debate by providing the most detailed and empirically sensitive content-based account we have, and their account might explain some of the data. However, we think that the current account has theoretical and empirical shortcomings that make it unfit to replace traditional approaches.

Our theoretical concern about D&P's proposal is that we are not given independent criteria for determining the explicit content of a subject's representation. Because their treatment of the evidence leans heavily on whether factuality is explicitly represented, what is especially urgent is a specification of independent criteria for determining whether factuality is explicitly represented.[1] For example, how does one determine whether a subject has a representation with the explicit content *it is a fact that the light is on* or just the content *the light is on*?

To press this problem a bit, consider D&P's claim that (bona fide) success on direct tests requires explicit representation of factuality (sect. 3.2, para. 6). Suppose you ask someone if Central Park is east or west of Lincoln Center, and the person gives the correct (and apparently bona fide) response, "east." What independent reason is there for thinking that he has an explicit representation of the form: *It is a fact that Central Park is east of Lincoln Center*? We see no empirical or nomological reason why the representation must explicitly include factuality. Of course, if we know that someone believes that *p*, then we would probably also agree that he believes that it is a fact that *p*. But typically this attributes only a "tacit" belief, that is, a belief that follows obviously from one's core beliefs and can be made explicit given the appropriate prompts (see, e.g., Fodor 1987c and Lycan 1988). So, if we believe that *p*, because we have a standing belief that *if* ψ, then it is a fact that ψ. But this is of no help for D&P's account, because such tacit beliefs do not count as explicitly represented.

Because it is difficult to determine when a person explicitly represents factuality, in evaluating D&P's proposal we think it is especially instructive to consult comparative evidence on monkeys and other animals that presumably lack the capacity to represent factuality explicitly. Comparative research indicates that there are striking similarities between monkeys and humans on implicit and explicit tasks.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

D&P maintain that, although the ventral and dorsal paths are independent components of the visual system, on their account the crucial difference is that the ventral path, unlike the dorsal path, represents factuality explicitly (sect. 4.1, para. 6; sect. 4.3, para. 5). However, because monkeys also have ventral and dorsal visual pathways, if monkeys can not explicitly represent factuality, then D&P's characterization of the ventral path in humans will not generalize to the ventral path in monkeys. Furthermore, monkeys with unilateral lesions to the striate cortex exhibit similar symptoms to human blindsight patients (sect 4.1, para. 7; see also Campion et al. "Structure, Function, and Consciousness in Residual Vision and Blindsight" *BBS* 6(3) 1983). Such monkeys, like humans with blindsight, can correctly classify lights and nonlights in their sighted half-fields, and they can detect and locate small objects and lights in their blind half-fields. Yet, again like humans with blindsight, when the monkeys are given the option to classify stimuli as lights and nonlights, they consistently classify the stimuli in their blind half-field as nonlights (e.g., Cowey & Stoerig 1995).[2] At least without further clarification, it is not obvious how D&P's account provides a principled explanation of these data. In particular, if blindsight in monkeys can not be attributed to a lack of explicit representation of factuality, presumably, blindsight in humans can not be so attributed either.

We also find important parallels between humans and monkeys in the research on memory impairments. Zola-Morgan & Squire (1990; see also Squire 1992) report striking correlations in performance on nonverbal memory tasks by humans and monkeys. Normal monkeys, like normal humans, pass declarative memory tasks. Monkeys with lesions to the hippocampus and medial temporal lobe, like amnesic patients, show impairments on crucial nondeclarative tasks. According to D&P, declarative knowledge requires explicit representation of factuality (sect. 3.3, para. 5), but if declarative memory is similar in monkeys and humans, it probably cannot be characterized as D&P suggest.

Traditional approaches to the implicit/explicit distinction offer a natural explanation of these findings. The comparative evidence indicates that damage to certain parts of the primate brain causes selective impairments on explicit processes, while leaving implicit processes intact. A traditional explanation of this is that damage to certain parts of the primate brain (e.g., the hippocampus) impairs the cognitive mechanism required for certain explicit processes (e.g., declarative memory). The challenge for a content-based approach like D&P's is to defend the idea that damage to certain parts of the primate brain causes selective impairment in the ability to represent certain contents and that this impairment applies only to certain processes. Without a response to this challenge, the traditional distinction between implicit and explicit processes seems to be indispensable.

Although we support the traditional cognitivist approach to the implicit/explicit distinction, we agree with D&P that this usage does not conform to the commonsense implicit/explicit distinction. But we think that this is not at all surprising. After all, the findings that motivate the cognitivist implicit/explicit distinction – for example, blindsight, selective memory impairment, the implicit understanding of false belief – are deeply counterintuitive. That is what makes them so fascinating.

NOTES

1. D&P maintain that if a thought is conscious, then it must explicitly represent factuality; but this does not really provide empirically useful criteria, for it depends on a controversial account of consciousness, and in any case, we have no independent criteria for identifying whether a thought is conscious.

2. Thanks to Lawrence Weiskrantz (personal communication) for directing us to this research.

Explicit to whom? Accessibility, representational homogeneity, and dissociable learning mechanisms

David C. Noelle

Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213. **noelle@cnbc.cmu.edu**
www.cnbc.cmu.edu/~noelle/

Abstract: Distinguishing explicit from implicit knowledge on the basis of the active representation of certain propositional attitudes fails to provide an explanation for dissociations in learning performance under implicit and explicit conditions. This suggests an account of implicit and explicit knowledge grounded in the presence of multiple learning mechanisms, and multiple brain systems more generally. A rough outline of a connectionist account of this kind is provided.

It is possible that our colloquial use of the terms "explicit" and "implicit" can help us understand the psychological mechanisms that subserve the memory and learning phenomena that have been labeled with these words. As Dienes & Perner's (D&P's) target article suggests, explicit mental processing may differ from implicit processing primarily in the amount and kind of actively represented knowledge involved. However, applying the common communicative notion of "explicit" to such processing introduces a new question: "Explicit to whom?" Different brain systems and mechanisms encode different kinds of information. The main difference between implicit and explicit knowledge might arise from differences in the mechanisms housing the knowledge, rather than from the presence or absence of a "factuality" propositional attitude. In other words, the telling feature that makes some knowledge explicit might not be *what* features are represented, but *who* (i.e., what brain systems) has access to the knowledge.

D&P's proposal appears to allow for the collocation of implicit and explicit knowledge in a single brain system, with both kinds of knowledge represented in the common currency of propositions and propositional attitudes. D&P do not deny that different brain systems may maintain different knowledge (citing, for example, evidence for dual pathways in the visual system), but what makes their theory distinctive is that it relates explicitness to the specific knowledge content of individual systems. This representational homogeneity poses problems for the theory, however, putting it at odds with observed performance differences between implicit and explicit learning tasks, and in an awkward position with regard to experiments investigating the transition between these kinds of knowledge.

Human performance on certain implicit learning tasks has been shown to be qualitatively different from performance on analogous explicit learning tasks (Shanks & St. John 1994). For example, learners attempting to produce explicit rules for controlling a complex system tend to acquire more simple, and often inferior, control strategies compared to those who approach the task without such a demand for linguistic description (Berry & Broadbent 1988). Similarly, categories learned incidentally, as part of performing another task, tend to have more of a "family resemblance" structure and less of a rule-guided structure than when an explicit effort is made to learn the categories (Brooks & Wood 1997). Research of this kind shows that knowledge acquired implicitly can be substantially different in form from that acquired explicitly. D&P's framework does not seem to capture this difference. Specifically, knowledge of an implicit kind differs in this framework from explicit knowledge only in its lack of an active representation of "factuality" and, perhaps "predication." Both implicit and explicit knowledge make use of the same vocabulary of predicates. Given this homogeneity of representation, it is hard to justify the observed differences in acquired knowledge without appealing to dissociable learning mechanisms. One may assert that such separate learning mechanisms exist without contradicting the proposed scheme of propositional attitudes, but the utility of those propositional attitudes is greatly reduced once multiple

learning systems are introduced. Instead of relying on a "factuality" tag, implicit and explicit knowledge may simply be that produced by, and embedded in, the corresponding kind of learning system.

Furthermore, if a common representational currency is used for implicit and explicit knowledge, then one might expect it to be relatively easy to transform implicit knowledge into explicit knowledge. All that is required is the introduction of the appropriate "factuality" attitude. For humans, however, such transformations are rarely easy. For example, in artificial grammar learning experiments, participants who implicitly acquire useful information about the regularities in some letter strings, and are then given explicit instruction in the systematic structure of the strings, often find it very difficult to integrate these two sources of information (Reber et al. 1980). Once again, this argues for separate representational spaces, or at least separate collections of predicates, for the two kinds of knowledge.

One alternative to D&P's proposal involves attending to the way information processing is distributed throughout the brain. Each neural system, receiving sustained input from some other system, may be seen as interpreting its input as some "explicit" representation of the knowledge provided by that other system. The proper use of this input typically requires some "implicit" knowledge concerning the processing to be performed – knowledge hidden in the circuitry of the receiving neural system. Thus, from the point of view of individual brain systems, the colloquial usage of the terms "implicit" and "explicit" makes sense. From the point of view of the animal as a whole, however, there is no explicit knowledge of most of this neural communication. Knowledge becomes truly explicit only when a member of a select class of brain systems might, given proper cuing, find that knowledge in its input. This class of systems might be thought of as primarily sensory in nature – allowing for the representation of the knowledge in some linguistic, auditory, visual, or other sensory code. Thus, under this alternative account, it is not the *content* of a representation that makes it explicit, but the *accessibility* of that knowledge as input to appropriate brain systems.

There is a natural connectionist framing of this alternative view. Connection weight values, which strongly influence behavior but are never directly visible to functionally adjacent brain systems, may embody only implicit knowledge. The activation states of processing elements, however, are communicated to other units, allowing them to "explicitly" encode information for downstream systems. When such patterns of activation can be made available to certain sensory systems, remaining stable for a substantial period of time (Mathis & Mozer 1995), they may be seen as representing explicit knowledge. This account is consistent with the differences between implicit and explicit learning performance, as separate mechanisms are present for the two styles of learning. Implicit learning may proceed via the modification of connection weights, whereas more explicit forms of learning may take place through the propagation of activation values (Noelle & Cottrell 1994). Perhaps most important, this connectionist account allows implicit and explicit knowledge to be distinguished without requiring an active marker of veridicality on most every representation in the brain.

What's really doing the work here? Knowledge representation or the Higher-Order Thought theory of consciousness?

Gerard O'Brien and Jonathan Opie

Department of Philosophy, University of Adelaide, Adelaide, South Australia 5005, Australia. **{gobrien; jopie}@arts.adelaide.edu.au**
arts.adelaide.edu.au/Philosophy/{gobrien; jopie}.htm

Abstract: Dienes & Perner offer us a theory of explicit and implicit knowledge that promises to systematise a large and diverse body of research in cognitive psychology. Their advertised strategy is to unpack this distinction in terms of explicit and implicit *representation.* But when one digs deeper one finds the "Higher-Order Thought" theory of consciousness doing much of the work. This reduces both the plausibility and usefulness of their account. We think their strategy is broadly correct, but that consensus on the explicit/implicit knowledge distinction is still a fair way off.

We are entirely sympathetic with Dienes & Perner's (D&P's) attempt to bring order to the confusing variety of ways in which the explicit/implicit distinction is employed in cognitive psychology. Moreover, we think an approach that unpacks explicit and implicit *knowledge* in terms of explicit and implicit *representation* (i.e., in terms of the different forms of information coding that a cognitive system can engage in) is well placed to "integrate and relate the often divergent uses of the implicit-explicit distinction in different research areas" (Abstract). D&P, however, give the impression that this systematisation can be achieved *solely* on the basis of a distinction between different forms of knowledge representation (see sect. 1, especially para. 9). If this is what they really intend, then we believe D&P's project is bound to fail.

To see why, consider the pioneering work of Reber (1967; 1989) on implicit learning. Reber demonstrated that subjects who memorise a set of strings generated by a finite-state grammar can subsequently perform above chance when judging the grammaticality of novel strings. Moreover, they can do so without any intention to learn a grammar, and with little or no capacity to articulate the relevant rules (although, see Dienes et al. 1991; Dulany et al. 1984; Perruchet & Pacteau 1990). The standard explanation is that during implicit learning subjects acquire implicit knowledge – unconscious representations (of either substrings, whole strings, or abstract rules) that guide subsequent behaviour. This kind of explanation does not seem to require any particular commitment to the form in which knowledge is represented. The distinction between conscious and unconscious representations is very much in the driver's seat. What makes knowledge implicit for Reber (and many others) is merely the fact that it is unconscious.

D&P are well aware of this emphasis on the conscious/unconscious distinction in the literature. In section 3.1 they explain how it relates to their representational reading of explicit/implicit knowledge. They suggest that most cases of conscious knowledge require, in addition to the explicit representation of knowledge contents, the explicit representation of an attitude to those contents, and of the self as bearer of that attitude. This is a version of the Higher-Order Thought (HOT) theory of consciousness: For knowledge with content X to be conscious, X must be the object of a second-order state (e.g., knowing that I know X; Carruthers 1996; Rosenthal 1986). If one adopts the HOT theory, it is clear why "most definitions of explicit knowledge involve consciousness; because it imposes the clearest, most extreme case of explicitness" (sect. 3.1, para. 5). Thus, according to D&P, knowledge is conscious when it is connected with a number of explicit representations, and it is unconscious (hence implicit) when it is connected with little, if any, explicit representation. Given this analysis it is not hard to integrate Reber's work with D&P's scheme. One has implicit knowledge of the rules of a grammar (in Reber's terms) if those rules are explicitly represented, and if there is *no* explicit representation of self or attitude.

This is where the trouble starts, however. To achieve this systematisation (integrating research that largely ignores issues of

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

representational form), D&P have had to introduce the HOT theory of consciousness. Theirs is a story about the forms of knowledge representation *combined with* a particular take on consciousness. Indeed, it looks as though the real burden in D&P's project is being carried by their preferred theory of consciousness. Without this theory there is simply no way for D&P to make sense of much research that invokes the distinction between explicit and implicit knowledge.

One might think that this is okay – we simply have to widen our conception of D&P's project. This is clearly an option, although it renders the initial statement of their theory somewhat misleading (in view of the fact that D&P claim that they are offering us a systematisation in terms of knowledge representation alone, when the HOT theory is actually doing much of the work). But the price one pays for this more generous conception is to reduce greatly the general appeal (and plausibility) of D&P's account. Although their story about the forms of knowledge representation attracts a fair degree of consensus (see, e.g., Cummins 1986; Dennett 1982; Pylyshyn 1984), the same cannot be said for the HOT theory itself. Whatever its virtues, this theory is by no means uncontroversial. We count ourselves among the many theorists who would deny that higher-order thoughts are either sufficient *or* necessary for consciousness (of either the phenomenal or access variety). Anyone in this position will thus fail to be satisfied by the proposed systematisation of the literature. The beauty of D&P's initial suggestion was that it promised to bring order out of chaos. But the chaos is bound to return as a result of some pretty entrenched debates about consciousness.

Let us finish by reiterating that we think the aim of the project is a good one, and that the general thrust of D&P's approach is inviting. Combining a principled story about knowledge representation with a theory that links consciousness and explicit representation will ultimately bear fruit, we believe, vis-à-vis the explicit/implicit knowledge distinction. Inserting our own favoured approach to consciousness – the Connectionist Vehicle Theory (CVT), which identifies phenomenal consciousness with the explicit representation of information in neurally realised PDP networks (O'Brien & Opie 1999) – one ends up with pretty similar-sounding conclusions. Explicit knowledge contents are *explicitly represented*, and hence are conscious. Implicit knowledge contents piggyback on what is explicitly represented, but are not themselves represented explicitly, hence they are unconscious. Of course, CVT has no more of a privileged status within consciousness research than the HOT theory. There just is no agreement on the best strategy for explaining consciousness. Consequently, we may have to live with the fact that real consensus on the distinction between implicit and explicit knowledge is still a fair way off.

A methodological requirement in the investigation of "knowledge"

Mark John O'Brien

School of Computer Science and Information Technology, University of Nottingham, Nottingham NG7 2RD, United Kingdom. **mark@cs.nott.ac.uk**

Abstract: The modernist and scientific juxtaposition of object and subject are inappropriate when investigating the nature of "knowledge." This commentary argues that the usual methodological dichotomy fails when it is applied to the domain of "knowledge." The two instead coalesce within the topic itself, demanding the most careful self-awareness.

When the young Wittgenstein of the Tractatus undertook his investigation into the status of propositions he did so through the use of propositions. This was no trick, for he knew well enough that both the subject and the object of such an investigation must not only be consistent but identical. Anything less would open him up to the criticism that his end and his means were incompatible, dealing a mortal blow to both. The failure of Dienes & Perner's (D&P's) target article to match the methodological purity and rigour of Wittgenstein's masterful approach condemns the work from the very outset.

For the purposes of elucidation, any work that proposes to generate knowledge on "knowledge" has at its disposal itself. The work *is* knowledge and thus any techniques, analysis, or indeed results, can be directed onto itself through an internal process of self-referentiality. It becomes its own exemplar. But what are we given here? "This is a cat"! Far better that the target article should have directed the light of its enquiry onto the knowledge encapsulated in the article itself; and in that vein we should see the analysis, and dismantling, of statements such as: "Knowledge is standardly analysed as propositional attitudes" (sect. 2.1.2). Now what is implicit here? *That* is a question worth answering.

This reflexive property of knowledge demands care and attention from anyone who dares to handle it; and such care and attention is missing from this article. Take, for example, the opening statement of section 2.1.3.: "It is evident"; if we are to allow the authors the luxury of rhetorical questions, then we might also ask "evident? In an article on knowledge?" An introduction to any claim in this manner ill-becomes an article on knowledge. And at second sight this becomes worse: "*It* is evident"? It? Reflection on the structure of the sentence and this use of a single, and unnecessary, auxiliary word destroys the very credibility that is sought by the authors. They have used a rhetorical device to weaken the opening of the sentence and thereby hope to lead the reader into an easy acceptance of the rest. If the claim is truly evident, then it should open the sentence and, correspondingly, the conclusion would be " . . . is evident." The point of this miniature textual analysis is to highlight D&P's use of rhetoric and linguistic devices to manipulate the knowledge that is presented within. Such devices occur throughout the text. This may have been premeditated. More charitably, the authors may not have even been aware of their own use of such literary tricks and sleights of hand, yet in this case they exhibit the very lack of reflexivity demanded by the subject.

This analysis and conclusion might seem harsh, but the general lack of care in the work extends beyond a purely stylistic and syntactic critique. To put it bluntly, the target article shows gay abandon with its use of terminology. A careful reading reveals a host of connotations and denotations of the word "fact," and these could have yielded rich insights if the authors had chosen to analyse it rather than the more banal word "bachelor." The clumsy and ultimately circular analysis of "bachelor" is amusing enough in its own right, but the article would have been exquisitely entertaining if the authors had chosen to tackle the word "fact" instead!

None of this is surprising. The approach adopted is essentially rationalist in form. The underlying meta-narrative is modernist, using scientific objectivism as the legitimating foundation. A more self-referential and ironic technique is demanded. The fatal flaw is the attempt to objectify the subjective. The blindness in this regard is obvious from the References: no Saussure. Nor are there any references to the developmental line that sprang from Saussure and eventually became the subject known as linguistics. One can understand why the authors have turned their face from that line and they are not the first to have done so. Piling mixed metaphor on mixed metaphor, such an approach would be to open up a can of worms that could not be nailed down; to the unwary it would be a slippery slope of regress.

So far this commentary has been more of a destructive critique, so perhaps some concluding positive suggestions are in order. To all readers of the target article, the key insight that will help in understanding it and its shortcomings is the need for awareness. To mix metaphor with reality: Read the article with your eyes open. As you read you should be aware of the article itself and deal with the text on its own terms. This is not easy and imposes a further requirement on the reader: self-awareness. One should be aware of one's own awareness; constantly validating oneself. Yet this still will not cut deep enough. Awareness of self-awareness demands the self-awareness of self-awareness and so on through a recursion

without end. As the hairdresser holds up the mirror to show you the back of your own head, and thereby shows a reflection of a reflection of a reflection of . . . you see yourself disappearing off to infinity at the speed of light.

And to those who seek to understand "knowledge," such behaviour should be used not only to read this article, but in a more important capacity, to guide and focus all your own inquiries. To return to where we started with Wittgenstein: The effect is to remove the duality of subject and object, the "I" and "not-I." As a Zen master might put it: Herein lies enlightenment.

One final important, yet tangential and elliptic, remark: This commentary demands its own commentary, which would reveal it to be as true as fiction, and as fraudulent as the poetry of Ern Malley.

What is special about "implicit" and "explicit"?

Geir Overskeid

Program in Economic and Organizational Psychology, Norwegian School of Management BI, Sofienberg, 0506 Oslo, Norway. **geir.overskeid@bi.no**
www.bi.no/users/fgl97015/index.htm

Abstract: Dienes & Perner present a very interesting analysis of two types of knowledge. It is not clear, however, that the words "implicit" and "explicit" are the best basis on which to build a theory of the two types of knowledge. One is also left uncertain as to whether this theory is the best way of ordering the greatest possible amount of relevant data in a way that yields the simplest account possible.

Starting in antiquity, many thinkers have addressed the difference between two types of human behavior – one based to a great extent on knowledge that can be formulated as maxims, rules, or hypotheses; the other governed to a greater degree by knowledge that is not, or cannot be verbalized. Both Socrates and Democritus discussed this distinction (see Overskeid 1995).

Since then, philosophers and psychologists have had, and still have, difficulties describing the phenomena related to these forms of knowledge. Because authors have not been able to agree on the nature of the two types of knowledge – indeed, they cannot agree that two clearly different types of knowledge exist (Cleeremans et al. 1998; Overskeid 1994a; Shanks & St. John 1994) – no terminological consensus has ever existed. This has led to a plethora of word-pairs purporting to describe the same, or closely related phenomena (see Overskeid 1994b).

Dienes & Perner's (D&P's) target article is a very interesting analysis of two types of knowledge, leading to predictions that map onto the empirical literature. It is not completely clear, however, why the authors chose the words "implicit" and "explicit" as the basis for their logical and conceptual analysis. Is there any reason to assume that these words are in some way better than the alternatives? One may wonder what might result from a similar analysis of concepts such as *imponderable* versus *ponderable* or *documentary evidence* (Wittgenstein 1958), *sensitive* versus *cogitative parts of our natures* (Hume 1969), *knowledge by acquaintance* versus *knowledge by description* (Russell 1961), or *experiential* versus *verbal knowing* (Hayes 1992). In later years, the words implicit/explicit have become popular in describing our two types of knowledge. They are, however, arbitrarily chosen from among many possible alternatives.

The role of theory in psychology has been debated (e.g., Overskeid 1999; Patton & Jackson 1991; Skinner 1950; Smedslund 1998). There is nevertheless some consensus that a primary function of a theory should be to order the greatest possible amount of relevant data in a way that yields the simplest possible account of the phenomena to be explained (see e.g., Chater 1997). If this is taken as a point of departure, doubts may be raised as to whether a framework based on a conceptual analysis of two words is the best starting point from which to further our understanding of the two types of knowledge. This question seems especially appropriate, insofar as no argument is offered in the target article as to why the words "implicit" and "explicit" describe the phenomena in question better than the many existing alternatives.

D&P stress the advantage their theory has in that "it is grounded in the ordinary use of terms 'implicit' and 'explicit'" (sect. 1), and they disapprove of other authors who have used the terms more or less as synonyms for "unconscious" and "conscious." Because this point is fundamental to their analysis, it deserves to be discussed. Based on what they call "the conceptual structure of the explicitly used words" (sect. 1), D&P state: "that someone is male and unmarried is a necessary supporting fact for the explicitly conveyed fact that he is a bachelor." Furthermore, the authors define a fact as something that is

> explicitly represented if there is an expression (mental or otherwise) whose meaning is just that fact; in other words, if there is an internal state whose function is to indicate that fact. Supporting facts that are not explicitly represented but must hold for the explicitly known fact to be known are *implicitly represented* (sect. 1, last para.).

Given the information above, we should ask: Where is the distinction between the fact that someone is a bachelor and the fact that he is an unmarried male? Because "unmarried male" is the unequivocal definition of "bachelor," we can hardly draw a meaningful distinction between the "explicit" fact that someone is a bachelor and the "implicit" fact that he is an unmarried male. In the absence of what D&P call implicitly represented "supporting facts," characterizing a person as a bachelor conveys no fact at all.

A conflict seems to exist between what D&P take to be the ordinary use of "explicit" and the dictionary's version. According to the Merriam-Webster Collegiate Dictionary (1994), "explicit" means "fully revealed or expressed without vagueness, implication, or ambiguity: leaving no question as to meaning or intent."

Now, if we follow D&P, stating just the fact that someone is a bachelor is an explicit statement, and something different from the implicit necessary supporting facts that he is male and unmarried. But how can calling someone a bachelor, isolated from the "implicitly" stated "necessary supporting facts" of the term, be an explicit statement – given that explicit statements, according to the dictionary, leave "no questions as to meaning or intent." Summing up, I think one might argue that (1) D&P's definitions are not fully based on ordinary language, and (2) one often cannot speak meaningfully of "explicitly conveyed facts" and implicit "supporting facts."

Based on their analysis of the meanings of "implicit" and "explicit," Dienes & Perner arrive at several conclusions supported by existing data. Though this is reassuring, one would feel even better as a reader if the authors had shown that their theory accounts for more facts in a simpler way than do other ways of understanding the same phenomena. Furthermore, all words may not be equal, but I miss the reasons why, in the words of Orwell, "implicit" and "explicit" are more equal than others. That is, how do we know that an even better theory could not have been built on one of the many other pairs of words available?

Knowledge by ignoring

Paul M. Pietroski[a] and Susan J. Dwyer[b]

[a]*Linguistics and Philosophy, University of Maryland, College Park, MD 20742-7615;* [b]*Department of Philosophy, University of Maryland-Baltimore County, Baltimore, MD 21250.* **pietro@wam.umd.edu**
dwyer@umbc.edu

Abstract: Some cases of implicit knowledge involve representations of (implicitly) known propositions, but this is not the only important type of implicit knowledge. Chomskian linguistics suggests another model of how humans can know more than is accessible to consciousness. Innate capacities to focus on a small range of possibilities, thereby ignoring many others, need not be grounded by inner representations of any possibilities ignored. This model may apply to many domains where human cognition "fills a gap" between stimuli and judgment.

Dienes & Perner (D&P) rightly distinguish grammatical rules from grammaticality judgments (sect. 4.4.2). They also go on to say that knowledge "of artificial grammars and of natural language may differ." *May?* Given that D&P focus on "rules . . . that the participant has *induced*" (our emphasis), the question seems to be whether there are interesting respects in which natural language acquisition is like artificial language learning. Chomsky (1981; 1986a) and others have argued persuasively that the most important principles governing natural language are not induced, but innately specified up to parametric variation. In any case, the Chomskian program constitutes an apparently successful branch of cognitive science where appeal to implicit knowledge has figured prominently, and been much discussed. We have argued elsewhere against representational theory of mind (RTM) construals of implicit linguistic knowledge (Dwyer & Pietroski 1996); even if we are right, however, this does not contradict D&P's account as applied to other domains. Indeed, we grant that some kinds of implicit knowledge are capturable by an RTM approach. But we doubt that RTM-models characterize "*the* most important type of implicit knowledge" (Abstract, our emphasis).

First, a quibble. D&P note that ordinary speakers "lack explicit knowledge of the grammar rules for English," adding that such speakers are "fully aware and have explicit knowledge" of "their ability to judge the grammaticality of English sentences" (sect. 4.4.2, para. 2), but not if "grammaticality" is construed (as D&P suggest) in terms of whether strings of words are well formed according to the grammatical rules. Grammaticality (in this sense) is a technical/theoretical notion that ordinary speakers do not grasp. And speakers often judge grammatical strings to be unacceptable; famous examples include: "The horse raced past the barn fell" and "The rat the cat the dog chased chased ate the cheese." That said, speakers seem to be fully aware and have explicit knowledge of whether they find a given string *acceptable;* and if Chomskian linguistics is on the right track, speakers' acceptability judgments are products of (*inter alia*) their implicit innate grammar, sometimes called their linguistic competence.

One might imagine that speakers are related to their grammar as follows: for every proposition P that *linguists* write down in a correct theory of the relevant language, speakers have internal (RTM-ish) representations of P; and these representations are (in part) causally responsible for the speakers' judgments that certain strings (viz., those that would be grammatical only if not-P) are unacceptable. Variants on this view have been in the literature for some time. (See Dwyer & Pietroski 1996 for a review.) But there is little to no independent evidence that the basic principles of grammar are so represented. And in our view, explanations given in linguistics are not hostage to particular representationalist assumptions.

For present purposes, let us assume that children face a severe poverty of stimulus problem in acquiring a natural language, and they fill the gap by exploiting cognitive resources available to them as part of their genetic endowment. On this view, children "solve" the poverty-of-stimulus problem by effectively ignoring an endless number of hypotheses compatible with (and perhaps even confirmed by) the available linguistic data. In this sense, the process of acquiring linguistic competence is *not* a process of responding to the world as an ideally rational (and open-minded) scientist would in the course of hypothesizing generalizations that cover the available data (and serve as the basis for predictions). The environment matters; but the now familiar story is that many stimuli are better viewed as causal triggers of cognitive resources, rather than data that bear rational relations to hypotheses, if only because the database seems too small. Put another way, the gap between available data and ensuing judgment would be too large for an *inductive* leap. One can reimpose the hypothesis-testing model, by saying that the child ignores many quite natural hypotheses, *because* the child comes equipped with explicitly represented background assumptions. Just as one might assign students the task of accommodating a certain range of data given some theoretical assumptions that are not to be questioned for purposes of the assignment, so one might think that nature sets children the task of acquiring an idiolect (compatible with available evidence) given certain assumptions that are fixed *a priori.* But we repeat: One need not assume that children have explicitly represented (on inner analogs of notebooks) principles of universal grammar. One might think instead that children never even consider the possibility that the language spoken around them fails to conform to those principles.

From this perspective, linguists seek to characterize the space of *humanly possible* languages – that is, the range of languages that children might actually acquire in natural environments – where this will be a (small) subset of the space of languages compatible with the noises produced in the child's neighborhood. But for children, one might say, there is no distinction between the class of languages and the class of humanly possible languages. It is not that children explicitly set aside some possible languages as irrelevant to their interests. It is that children have, from the ideal theorist's point of view, an impoverished concept of language; and lucky for them, because this innate capacity to ignore is precisely what children need to acquire *some* language on the basis of available cues.

Taken out of context, it sounds odd to say that capacities to *ignore* are sources of (even implicit) knowledge. But there is a point to talking about the relation that humans bear to the principles of universal grammar. This relation has as much title, both historically and theoretically, to the label "implicit knowledge" as any other. And this relation may well be rooted (not in inner representations whose contents are unavailable to consciousness, but rather) in an innate capacity to ignore. Again, we have nothing against RTM-ish notions of implicit knowledge. But the capacity to *not* consider possibilities is also a valuable (and probably crucial) cognitive resource. And it would be unsurprising if this kind of implicit knowledge figured in other domains where humans tacitly "fill a gap" between stimuli and judgment.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

A developmental theory of implicit and explicit knowledge?

Diane Poulin-Dubois and David H. Rakison

Centre for Research in Human Development, Department of Psychology, Concordia University, Montreal, Quebec, Canada, H4B 1R6.
{dpoulin; rakison}@vax2.concordia.ca

Abstract: Early childhood is characterized by many cognitive developmentalists as a period of considerable change with respect to representational format. Dienes & Perner present a potentially viable theory for the stages involved in the increasingly explicit representation of knowledge. However, in our view they fail to map their multi-level system of explicitness onto cognitive developmental changes that occur in the first years of life. Specifically, we question the theory's heuristic value when applied to the development of early mind reading and categorization. We conclude that the authors fail to present evidence that dispels the view that knowledge change in these areas is dichotomous.

Dienes & Perner (D&P) are to be praised for their attempt to enrich the simple implicit-explicit dichotomy that has prevailed in cognitive science over the last few decades. Early cognitive development involves substantial changes in the content, structure, and processing of information; it might accordingly be considered the perfect domain for D&P's theory. A critical issue for us, however, is whether their theory offers any new insight or heuristic for the study of cognitive development in infancy and early childhood. Past and present approaches to cognitive development have emphasized one or more of these changes. Those adopting a Piagetian and neo-Piagetian approach, for example, stress the development of structure, whereas information-processing theorists – including connectionists – focus on developmental changes in the processing and structure of information (Klahr & MacWhinney 1998). Karmiloff-Smith's (1992) [see also *BBS* multiple book review of Karmiloff-Smith's *Beyond Modularity BBS* 17(4) 1994] theory of representational redescription successfully integrates different aspects of these approaches in an explanation of cognitive change. Her theory offers a more direct conceptualization of cognitive growth, in which the implicit-explicit issue occupies center stage. Change in explicitness is not simply a peripheral characteristic of cognitive development; it differentiates one level of knowledge from another.

In applying their theory to the developmental literature, D&P draw a parallel between their four-level hierarchy of explicitness and the three-level system posited by Karmiloff-Smith (1992). However, despite the claim that their levels of explicitness yield a plausible correspondence" (sect. 4.3) to that developed by Karmiloff-Smith, in our view D&P fail to map the different levels of their hierarchical system onto a distinct development phase or stage. For example, the authors claim that level-I in Karmiloff-Smith's theory leaves predication implicit; however, they fail to identify a level corresponding to a proposition in which the individual is implicit but the property remains explicit. This shortcoming is most apparent in D&P's use of evidence from the theory of mind literature. For example, they draw a simple distinction between implicit understanding based on the abstraction of situational regularities and explicit understanding based on the causal understanding of belief formation. More specifically, in their account of changes in children's responses in false-belief tasks (e.g., Clements & Perner 1994), D&P contrast young children's visual orienting responses with older children's verbal explanations. Does the visual orienting response correspond to a proposition where predication has become explicit – that is, Karmiloff-Smith's (1992) level-E1 – or to implicit understanding similar to Karmiloff-Smith's level-I? In the absence of a clear answer to this question, the theory fails to provide a nondichotomous developmental theory and therefore has little heuristic value for researchers in this domain of knowledge.

D&P's proposal becomes more problematic when considering the earlier stages of mind reading, and in particular, infants' and toddlers' explicit reasoning about desire and emotion. Wellman and Woolley (1990), for example, found that children as young as three are able to provide a verbal explanation for someone's emotional state as a function of the fulfillment of that person's desires. In light of this evidence, we believe that D&P need to incorporate levels of reasoning about desire and belief to have a valid developmental theory. The problem may lie less with the theory itself and more with its application to a task – namely, the classic false-belief task – that is now considered too limited to assess the full development of mind-reading abilities (see Lewis & Mitchell 1994; Poulin-Dubois 1999).

In our view, the lack of heuristic value in D&P's theory is highlighted when it is applied to infant cognitive development. In section 4.3, the authors do not extend their theory to changes in knowledge that occur prior to the third year of life. However, there is currently some controversy in the infancy literature about the representational format of early knowledge and how it changes. Much of the debate focuses on whether there are two parallel representational systems – a procedural-type system and a declarative-type system – or whether infancy is instead characterized throughout by implicit knowledge (Mandler 1998). Nowhere is this debate better illustrated than in the area of infant categorization where the question of whether knowledge is perceptual or conceptual is hotly disputed. Mandler (1992; 1998), for example, has argued that by the end of the first year infants possess both implicit, procedural knowledge in the form of sensorimotor representations and declarative, explicit knowledge in the form of conceptual representation. In contrast, others (e.g., Poulin-Dubois et al. 1999; Rakison & Butterworth 1998) have argued that infants begin to acquire conceptual knowledge toward the end of the second year, and even then perceptual information is still prime in category membership judgments. What both views share is the simplistic notion of knowledge as implicit (or perceptual) or explicit (or conceptual). This seems like the perfect arena to apply D&P's theory. The theory should be able to provide a more detailed account of the transition from implicitness to explicitness within and between perceptual and conceptual knowledge. For example, how does implicit knowledge acquired in infancy through the perceptual array (as in sensitivity to biomechanical motion; Berthental 1993) become explicit conceptual knowledge concerning the motion characteristics of different kinds of objects (e.g., Gelman et al. 1995)? It is difficult to see how their theory could be applied to this important aspect of cognitive change.

In conclusion, we acknowledge the potential contribution of D&P's theory to changes in representational knowledge. One of its potentially fruitful applications to cognitive development might be to account for the dramatic changes in memory – for example, from procedural to semantic to episodic memory – in the first four years of life. D&P's ideas about the gradual explicitness of self or attitude might help explain the late emergence of autobiographical memory. Nevertheless, as it stands, it is difficult to apply their theory to developmental changes in young children's representations. D&P should expand their theory to allow such an application and to direct researchers to other developmental areas where it might be relevant.

ACKNOWLEDGMENT

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to the first author. We would like to point out that the order of authorship was determined by the toss of a coin.

Applying the implicit-explicit distinction to development in children

Ted Ruffman

Department of Experimental Psychology, University of Sussex, Brighton, East Sussex BN1 9QG, United Kingdom. **tedr@epunix.sussex.ac.uk**

Abstract: This commentary focuses on how Dienes & Perner's (D&P's) claims relate to aspects of development. First, I discuss recent research that supports D&P's claim that anticipatory looking in a false belief task is guided by implicit knowledge. Second, I argue that implicit knowledge may be based on exposure to regularities in the world as D&P argue, but equally, it may sometimes be based on theories that conflict with real world regularities. Third, I discuss Munakata et al.'s notion of graded representations as an alternative to the implicit-explicit distinction in explaining dissociations in infancy.

Dienes & Perner (D&P) have put together a very interesting and coherent framework for understanding the implicit-explicit distinction. One great strength of their proposal is that it generates testable predictions. This commentary highlights some of the developmental issues that arise from their ideas.

D&P argue that the anticipatory looking that precedes correct verbal performance in a false belief task is likely to be implicit. Yet they acknowledge that their arguments are based largely on intuition because no direct tests of whether the knowledge is conscious have been carried out. Recently, we tested these ideas by asking children to "bet" counters on where they thought a story character (Ed) would look for an object (Ruffman et al. 1998). In the False Belief task Ed placed the object in a left hand location and did not see it moved to a right hand location. In the True Belief task Ed saw the transfer. We replicated Clements and Perner's (1994) finding that children looked to the correct location when anticipating Ed's return before they answered the verbal question correctly. This could have been because eye movements indexed unconscious knowledge or because children were conscious that Ed *might* look in the left location but were not very confident. Our findings supported the first interpretation. Children who showed appropriate eye movements but incorrect verbal performance bet with great certainty that Ed would return to the right location (consistent with their verbal answer). Important to note, betting was a sensitive measure of even slight variations in certainty. When shown a bag containing 10 red marbles and 0 green ones, children bet with great certainty that a marble chosen from the bag would be red. Yet when the bag contained 9 red marbles and 1 green one, certainty that it would be red dropped off dramatically. So we have empirical evidence consistent with the idea that eye movements in a false belief task index truly unconscious knowledge.

The status of some of their other suggestions, though equally interesting, are perhaps less certain. Are eye movements really based merely on observed regularities as opposed to a theory? Previously, I found that 4- and 5-year-old children make a striking error (Ruffman 1996). Children were shown a round dish that held red and green sweets, and a doll was given the ambiguous message that "a sweet" from the round dish would be placed in a box. However, only the child saw as a red sweet was placed in the box. Children were asked what colour the doll would think the sweet was and over two trials they tended to claim that the doll would think there was a green sweet in the box. In this experiment children responded verbally (explicitly) to the experimenter's question. We are now running an implicit version of this task in which green sweets go in a green box and red sweets in a red box. We test whether children's eye movements indicate they also expect the doll to look for the sweet in the green box or the red box. The important point about this task is that there are no observed regularities where people who receive ambiguous messages consistently do the wrong thing. They should do the right thing (i.e., look in the red box) equally often. At the same time, there are various reasons why a bias to say the doll will get it wrong (i.e., look in the green box), suggests that children are using a theory. The results are not yet in, but the point of this example is that they could go either way. This means that claims about implicit knowledge being based on observed regularities might be correct or incorrect.

Finally, there are some aspects of development such as infancy that D&P do not specifically address. For example, even after 15-sec delays, 8- to 12-month-old infants look longer when an object is hidden in one location but retrieved from another. This seems to indicate surprise and to show some knowledge of the object's location. Nevertheless, the infants reach to the wrong location for the object (Ahmed & Ruffman 1998). D&P do not address dissociations between reaching and looking but their analysis provides some scope for doing so. In their section on visual perception, D&P suggest that pointing to an object may not require the object's property (location) to be explicit but leave it open as a possibility. The reaching-looking dissociation can be understood if reaching requires the location to be explicitly represented, whereas looking does not. Reaching would require explicitness because the delay imposed on the infant necessitates deliberation about the object's location. Looking would not require explicit knowledge because it requires nothing in the way of a declarative act that states its case. If this explanation is correct, one contradiction must be resolved. This is that infants show the looking effect even after delays of 15 sec, whereas D&P argue that representations that do not mark factuality (including reaching) fade after only a few seconds. One solution might be that lingering memories revealed through reaching do not necessitate a distinction between past and present, and hence do not mark factuality.

Another interpretation of the reaching-looking dissociation (Munakata et al. 1997) is in terms of graded representations (GR). GR holds that knowledge is gradually strengthened in development (analogous to connection strength), and reaching requires stronger representations. Munakata et al. make no reference to the implicit-explicit distinction, but GR allows that reaching and looking could be equally implicit. There are limitations to this view when applied to the findings for false belief understanding because children's betting seems to necessitate an appeal to different degrees of consciousness. Yet it seems a possible explanation of many dissociations in infancy. The challenge is to specify when the implicit-explicit view seems the best explanation of a finding and when GR seem best, or at least how to adjudicate between the two possible explanations.

Some costs of over-assimilating data to the implicit/explicit distinction

Mark A. Sabbagh[a] and Benjamin A. Clegg[b]

[a]*Developmental Psychology, University of Michigan, Ann Arbor, MI 48109-1109;* [b]*Department of Psychology, University of Surrey, Guildford, GU2 5HX, Surrey, England.* **sabbagh@umich.edu** **b.clegg@surrey.ac.uk**

Abstract: We applaud Dienes & Perner's efforts while raising some concerns regarding their assimilation of diverse data into a unifying framework. Some of the findings need not fit the framework they suggest. It is also not always clear what, above logico-semantic consistency, assimilation adds to the data that do fit their framework. These concerns are highlighted with reference to their arguments regarding the developmental data and the neuropsychological data, respectively.

Dienes & Perner (D&P) render an excellent service by noting that the notion of "implicit knowledge" as it has been bandied about through various subfields of cognitive psychology has become cloudy. As researchers who have grappled with understanding each other's use of the term "implicit," we applaud this effort to develop universal terminology. D&P's approach is thought-provoking and lays important groundwork for future research and theorizing on these issues.

Nonetheless, the attempt to broaden the scope of the term "implicit" does not come without costs. We offer two major concerns

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

regarding the effort to assimilate a larger set of data. The first is that it has led to the inclusion of phenomena that need not be thought of in implicit/explicit terms. The second is that questions remain regarding what is ultimately gained by assimilating previous findings involving the implicit/explicit distinction into their revised, more complex framework. We highlight these concerns from the standpoint of developmental and neuropsychological data, respectively.

D&P propose that there is an implicit/explicit developmental shift in children's conceptual development, suggesting that Clements and Perner (1994; 1996) have shown that when children respond in a nondeclarative mode (either with eye movements or a nondeclarative action) they produce evidence for an understanding of false belief, well before a similar understanding is available to declarative response modalities. However, there is some question as to whether this fits neatly into D&P's framework. In addition to making the response modalities nondeclarative, Clements and Perner have also made them "non-canonical," or atypical with respect to how one typically provides information in experimental or play settings. Carlson et al. (1998) found that 3-year-olds were better at deceiving others (i.e., deliberately indicating to someone that an object was in one place when really it was in another) when providing the deceptive information required pointing via a cardboard arrow as opposed to the more typical gestural pointing. Both of these response modalities are declarative; yet, when the non-canonical modality was employed, children demonstrated earlier competence. Carlson et al. suggest that by making the response type noncanonical, children were freed from their initial predisposition to provide the class of information they provide in canonical declarative actions (i.e., true information). Similarly, it could be that Clements and Perner's studies reveal correct responses from younger children simply because they have made the response modality noncanonical, and not because they have tapped predicate-implicit knowledge structures.

We believe that what might otherwise seem a relatively minor quibble with a particular data set is noteworthy because it may relate to a somewhat larger problem in applying D&P's framework. In offering a solution in which a variety of psychological phenomena are related in logical terms, D&P have shifted the focus away from distinctions that may be equally interesting, and perhaps more compelling. In the developmental literature, there has been a long-standing distinction between competence and performance (e.g., Chomsky 1986b). (Competence is the supporting knowledge base and performance is the capacity for expression of that knowledge base.) Although the competence/performance distinction could be assimilated into the implicit/explicit distinction, there is an important difference. As D&P apply their framework to the developmental literature, the implicit/explicit distinction hinges on the concomitant nondeclarative (predication-implicit) versus declarative (predication-explicit) distinction. By contrast, a competence/performance distinction allows for the possibility that predication may be explicit in the representation, but expression of the predicate might be hampered by resource limitations in the recruitment of executive functions. This is likely to be true when the expression of "Fb" goes against established conventions, as it does in false belief or deception tasks.

The loss of former distinctions would be welcome if a new taxonomy brought us to a deeper understanding of the psychological similarities shared by phenomena previously considered divergent. However, we question whether D&P have actually achieved this goal. Many of the proposals that invoke an implicit/explicit distinction are founded on attempts to make sense of apparent dissociations noted in the neuropsychological literature, be it the evidence for two pathways for visual processing (Ungerleider & Mishkin 1982) or the role of the hippocampus in memory (Squire 1987). Although these findings seem to fit into D&P's framework, this assimilation provides little insight into how their framework is testably different from other related distinctions. For example, might apparently diverse phenomena that share levels of implicitness be expected to show: similar patterns of development, similar patterns of breakdown in the event of brain trauma, similar activation patterns in neuroimaging studies, or similar neural mechanisms in operation? If D&P's ultimate aim is to highlight functional similarities, a more rigorous analysis of the neuropsychological data has important implications.

There is, of course, a question as to whether the goal is to highlight functional similarities. D&P conclude with the caveat that their "analysis of the meaning of implicit is in itself neutral on the question of whether different systems are responsible for producing knowledge of different degrees of implicitness" (sect. 5, para. 7). If this is the case, many interesting problems that might be addressed within an integrative framework seem to recur. Instead of having myriad implicit/explicit distinctions, are we left with myriad "predication implicit/explicit" distinctions? This is not necessarily a bad thing – their analysis provides an enriched grasp of what we mean by invoking the implicit/explicit distinction, and good guidelines for consistent usage. However, if consistency is all we are after, it could just as easily be maintained by exclusive domain-specific terminology as by expanding under an umbrella term.

In sum, it troubles us a little that in support of their framework, D&P may have incorporated evidence that need not be thought of in terms of the implicit/explicit distinction and that they have not provided enough of a rationale for assimilating data that do not in themselves suggest a more complex implicit/explicit framework. Despite these reservations we certainly share the intuition that use of the term "implicit" is not coincidental; undoubtedly, the natural meaning of the term underlies its initial adoption in a variety of research domains. Perhaps this is indeed the result of something ubiquitous in each of the situations. The question is worth asking and the proposal on offer is a bold step toward a lofty goal.

Representation and knowledge are not the same thing

Leslie Smith

Department of Educational Research, Lancaster University, Lancaster LA1 4YL, United Kingdom. **l.smith@lancaster.ac.uk**

Abstract: Two standard epistemological accounts are conflated in Dienes & Perner's account of knowledge, and this conflation requires the rejection of their four conditions of knowledge. Because their four metarepresentations applied to the explicit-implicit distinction are paired with these conditions, it follows by *modus tollens* that if the latter are inadequate, then so are the former. Quite simply, their account misses the link between true reasoning and knowledge.

Dienes & Perner (D&P) deserve credit for their dual focus in setting out an epistemological account of knowledge for direct application in current psychology. This dual focus is rare. Even so, their account of what it is for an epistemic system to know some fact is vulnerable to a counterargument in three parts. First, D&P conflate two standard epistemological accounts. Second, this conflation runs through their four conditions (o)–(iii) of knowledge. Third, D&P miss the link between true reasoning and knowledge.

The two available accounts of knowledge are the foundational and causal accounts. Under the foundational account, to know that p requires three criteria to be met: (a) p is true proposition; (b) the knower believes p; and (c) p has a justification available to the knower. Under the causal account, to know that p requires condition (a) alone to be met along with two further conditions: (d) p has a reliable, causal generation in the knower's mind; and (e) there is no other causal process responsible for p's generation in the knower's mind. The relevance of the accounts for psychology is discussed elsewhere (Smith 1992; 1993; sect. 13). But they are distinct accounts. They share condition (a) with conditions (b) and (c) present in the foundational account and absent from the

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

causal account, and conversely for conditions (d) and (e). Yet these two accounts are conflated in D&P's own account with disastrous consequences for their four conditions (o)–(iii) in section 2.1.2.

Condition (o). The unit of analysis used by D&P is a representation *R.* This is in contrast to the unit of analysis in both available accounts, namely, a proposition *p.* According to Frege (1979, p. 129), only a thought (*Gedanke*) can be true or false. Indeed, Frege (1977, p. 2) specifically denied that a representation (*Vorstellung*) can have a truth-value. Russell (1964, p. xix) followed suit, characterising a thought as a proposition. D&P have made a category-mistake in treating representation as having a truth-value. There can be an actual representation of a false proposition, such as *Austria is a state in Australia* (call this *p*). What is false is *p,* not its mental representation. Of course, if this actually is what is being represented, this fact, too, can be formulated as the true proposition *What I am now representing in mind is p.* Propositions can be true or false, but representations can be neither. For reasons given elsewhere (Smith 1998), representational subjectivity is devoid of truth-value, amounting to a "blind play of representations, less even than a dream" (Kant 1933, A112).

Condition (i). D&P's second condition states that "*R* is accurate (true)." Which is it? There can be degrees of accuracy, but not truth (Frege 1977, p. 3; 1979, p. 195). Indeed, a "false" representation (cf. "false memory") can itself be accurate. In a football match, a striker M scores the winning goal. Was this a brilliant header or did M handle the ball? M attests the accuracy of his recollection: a header and no hands. But a TV video replay shows that M did handle the ball. If there can be an accurate recollection of a false state of affairs, this means that accuracy and truth are not the same thing.

Condition (ii). This condition requires a judgment to be made. But a judgment is not just a correct response. Nor is it merely the recognition of truth. Rather, a judgment is the recognition of something as true (Frege 1979, p. 7). This is usually accomplished "by going back to truths that have been recognised already" (Frege 1979, p. 175). In short, any judgment requires a justification whereby the knower links a judgment to other judgments. In the best case, recognising the truth of a judgment by reference to a justification amounts to making a logical link. That is because logic is the science of truth (Frege 1979, p. 128). In fact, children's reasoning may not be like this because of their non-differentiation of empirical and logical relationships (Piaget 1923; 1968). In short, a judgment requires a justification, and this leads directly to the conflation that undermines condition (iii).

Condition (iii). This condition requires a representation to have a reliable causal origin "which when made explicit serves to *justify* the claim to knowledge." As such, this condition conflates the foundational account, which includes criterion (c) above but not (d), with the causal account, which includes criterion (d) above but not (c). There is a further conflation of the causal and the logical. Frege (1979, pp. 2–3) pointed out that thinking always has causal antecedents. It is for empirical psychology to identify these antecedents. But this admission leaves untouched the logical reasons that tether one judgment to another. If the capacity to make judgments develops during ontogenesis, this entails the development of the capacity to give reasons for judgments. It is one thing for adult experimenters to attribute these capacities to children; it is quite something else to ascertain which reasons children actually give, whether psychological or logical (Smith 1999b).

In general, two individuals cannot have the self-same representation. So they cannot have the self-same knowledge. Yet the Pythagorean theorem is a public object of knowledge, unlike *my idea of* the Pythagorean theorem to which I alone have access (Frege 1977, p. 16). D&P's representational account provides an inadequate analysis of objective (true-false) and intersubjective (self-identical) knowledge, which is publicly available to us all (Smith 1999c). In consequence, D&P's specification of the difference between implicit-explicit knowledge collapses. That is because each of their metarepresentations (0)–(3) is paired with a corresponding condition in section 2.1.2. Metarepresentations (0)–(3) entail their conditions (o)–(iii), so it follows by *modus tollens* that if the latter are inadequate, then so are the former.

There is another way, which is to take seriously questions raised long ago by Piaget (1923) about the "study of true reasoning." How does the notion of truth develop in the child's mind (Piaget 1995, p. 184)? How do empirical reasons constructed in time develop into atemporal necessities, which are true throughout time (Piaget 1986)? One basis for a psychological answer to these questions can be found in the epistemologies of Frege and Piaget (Smith 1998; 1999a).

Implicit versus explicit: An ACT-R learning perspective

Niels A. Taatgen

Department of Cognitive Science and Engineering,University of Groningen 9712 TS Groningen, The Netherlands. **n.a.taatgen@bcn.rug.nl**
tcw2.ppsw.rug.nl/~niels

Abstract: Dienes & Perner propose a theory of implicit and explicit knowledge that is not entirely complete. It does not address many of the empirical issues, nor does it explain the difference between implicit and explicit learning. It does, however, provide a possible unified explanation, as opposed to the more binary theories like the systems and the processing theories of implicit and explicit memory. Furthermore, it is consistent with a theory in which implicit learning is viewed as based on the mechanisms of the cognitive architecture, and explicit learning as strategies that exploit these mechanisms.

The distinction between implicit and explicit knowledge, memory, and learning is used with many slightly different meanings in the cognitive sciences. Dienes & Perner (D&P) show how these different meanings can be captured by a system in which the natural language meaning of implicit and explicit is used. In a sense, the title of the target article, "A theory of implicit and explicit knowledge," is misleading. It is rather a theory of how scientists use the terms implicit and explicit knowledge. A real theory of implicit and explicit knowledge should first answer the question of whether it is useful to have the distinction at all (cf. Newell 1973). The interesting point that the D&P theory supports, but on which it fails to capitalize, is that the distinction is not so fundamental after all.

It is useful to examine theories that stipulate that the difference is fundamental. According to the systems theory, for example, implicit and explicit knowledge are stored in separate memory systems (Squire & Knowlton 1995). The processing theory (Roediger 1990), on the other hand, supposes that different processes are used to store and retrieve information. The common property of both theories is that they propose fundamentally different mechanisms in the information processing architecture for implicit knowledge on the one hand, and explicit knowledge on the other hand. So why are these distinctions made? They are needed to explain certain empirical phenomena. Most of these phenomena are so-called dissociations that show that implicit knowledge is much more robust than explicit knowledge. Whereas implicit knowledge persists over a longer time period, explicit knowledge is quickly forgotten (e.g., Tulving et al. 1982). Amnesics have lost their ability to retain explicit knowledge, although their implicit memory is intact (Warrington & Weiskrantz 1970). Individual differences in implicit learning are small, whether they are the result of age or intelligence, whereas individual differences in explicit learning are large (e.g., McGeorge et al. 1997). These empirical results are part of the reason we can talk about implicit versus explicit knowledge instead of just conscious and unconscious knowledge. These are the data that need to be explained by a theory of implicit and explicit knowledge. Neither the systems nor the processing theory is entirely satisfactory: they propose separate mechanisms to explain the distinction. A unified account would be preferable.

Unfortunately, D&P's theory offers only some starting points

for a unified explanation. In my view, a proper account of implicit and explicit knowledge should start with a theory of implicit and explicit learning, because the explicitness of knowledge, as D&P indicate, depends on the content in which it is acquired, and whether or not this context is retained. A useful approach is to view the distinction using the ACT-R architecture (Anderson & Lebiere 1998). ACT-R is a cognitive theory implemented in a simulation system that can be used to model performance and learning on individual tasks. The architecture encompasses several learning mechanisms. For example, one of the learning mechanisms keeps track of how often certain information in memory is needed, and adjusts certain activation parameters accordingly. The learning mechanisms, however, are all quite primitive: there is no mechanism that performs analogies or other complex forms of reasoning (as opposed to its predecessor, ACT*). To perform complex reasoning, the system needs additional knowledge, which has to be applied in a goal-drive fashion. So to gain new knowledge by using analogy, an explicit analogy goal has to be posed, and procedural knowledge needs to be supplied to retrieve an example and find the appropriate mappings.

The learning mechanisms of the architecture take care of the fact that the results are stored and evaluated for their usefulness. Implicit learning seems to correspond very well with the learning mechanisms in the architecture. These mechanisms are always at work and are not directly related to the current goals of the system. Because they are not tied to the goals of the system, they are not directly available to consciousness. Explicit learning, on the other hand, is tied to goals, and is dependent on procedural knowledge. This means that a certain type of explicit learning is only possible if the proper knowledge is available. This also explains why individual differences in explicit learning are so large. It also implies awareness, because the acquired knowledge is associated with a learning goal. I have shown (Lebiere et al. 1998; Taatgen 1999), that this way of looking at the distinction allows explanations for several of the implicit learning phenomena. This theory also avoids a binary distinction between implicit and explicit learning: explicit learning is just a clever way of processing information so that the implicit learning mechanisms pick up the right information. In a sense, all learning is implicit learning.

At this point it is useful to compare this account to D&P's theory. According to D&P, information is more explicit as more information about its justification and attitude is available. In the ACT-R account, information is explicit if there is a learning goal associated with it. This learning goal may serve as a source of justification, because it contains information on the success of the goal, and may also point to other contextual information, like attitudes.

Automatic processing results in conscious representations

Joseph Tzelgov, Dana Ganor, and Vered Yehene

Department of Behavioral Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel 84105. **{Tzelgov; Yehene}@bgumail.bgu.ac.il**
www.bgu.ac.il/beh/yossi.html

Abstract: We apply Dienes & Perner's (D&P's) framework to the automatic/nonautomatic processing contrast. Our analysis leads to the conclusion that automatic and nonautomatic processing result in representations that have explicit results. We propose equating consciousness with explicitness of aspects rather than with full explicitness as defined by D&P.

Dienes & Perner (D&P) provide a detailed analysis of representations. They define three components of a representation: C3 – the holder of it (self), C2 – the attitude, and C1 – the content. C1 is further divided into specific aspects: an object, its property (or properties), and a proposition predicating these two. D&P propose that if a given component C1 is explicit, all "lower" components have to be explicit.

D&P conceptualize consciousness in terms of higher order thought theory (see sect. 3.1). This equates consciousness with full explicitness up to C3 and focuses on the monitoring aspect of consciousness (D&P n. 11). D&P (sect. 3.4) propose that monitoring defines "willed action" or "nonautomatic processing."[1]

We agree with this conceptualization. We also agree that automatic processing, as the complement of nonautomatic processing, is defined by the absence of monitoring because processing without monitoring is the single feature common to all automatic processes (Bargh 1992). Such processing is best indicated when it takes place, although it is not part of the task requirement. Hence Stroop-like phenomena may serve as indications of automaticity (Tzelgov 1997).

Processing without monitoring does not require full explicitness. This leads to the conclusion that automatic processing is based on unconscious representations. In what follows we challenge this conclusion. Consider the process of reading individual words. The Stroop effect indicates that this may be automatic in the sense of "processing without monitoring." According to D&P's analysis, people showing the Stroop effect are unconscious of the representations of the words involved. But are they? It could be argued that under such conditions automatically processed words are not consciously perceived. The results of Marcel (1983a), which point in this direction are very hard to replicate (Holender 1986). Tzelgov et al. (1997) have shown that in tachistoscopic presentations, the Stroop effect is constrained to trials in which subjects are able to report the word and to subjects who show above chance recognition memory of the color words. These results imply that automatic processing may be based on conscious representations that are available to conscious (explicit) memory, and thus are, at least in part, content-explicit. Given the pattern of results obtained by Tzelgov et al. (1997), one could wonder what characterizes the representation resulting from automatic processing. It may be that such processing does not result in a proposition that predicates the automatically processed property with the object to which it belongs (Dulany 1991). Preliminary results obtained in our lab by D. G. suggest that this may be true. In a recognition test performed after a test where subjects were asked to report the color-presented words, the number of false positives was significantly higher for synonyms of the presented words than for control words. This is consistent with the idea that a specific property (i.e., meaning) of the words was explicitly represented, but no proposition relating this property to a specific visual form (the word presented) was generated.

D&P contrast automatic processing with willed action. We prefer the term "nonautomatic processing" to "willed actions" because frequently willed actions have automatic components. Vallacher and Wegner (1987) have pointed out that any action can be represented at many levels and that people tend to represent it at the highest level possible, which acts as the highest level "source schema" (Norman & Shallice 1980). Action is monitored at this level. The actions at the subordinate levels are performed automatically if a person is able to do so (Vallacher & Wegner 1987). Suppose that a person who is able to read automatically, as indicated by a Stroop effect, is asked to read a sentence, "The girl looked at the blue sky," and to decide whether it is true. In this case the sentence is read for meaning. The read sentence, being the monitored entity, is represented explicitly up to the level of C3. However, the individual words of that sentence are read automatically because they are backgrounded to the extraction of the meaning of the sentence (Jacoby et al. 1992). Consequently, those properties of the words that are relevant to the action specified by the preponent schema (e.g., the meaning of the word "blue") are represented explicitly and result in a proposition(s) that reflect their relevance to the extracted meaning of the sentence ("The sky was blue"). However, no propositions predicating meaning to specific visual patterns (e.g., "this visual pattern means 'blue'") are generated. Consistent with this analysis, it has been suggested that in sentences read for meaning only the gist of words is retained (e.g., Sachs 1967). Unpublished findings of D. G. provide direct

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

evidence supporting our argument: Synonyms of words in sentences read for meaning are more frequently falsely recognized as appearing in the sentences read, than control words. Thus it seems that the meaning or gist of words read automatically (either intentionally as a part of a sentence or autonomously, as in the Stroop task) is represented explicitly. We believe that under such conditions the reader is conscious of the meaning of the words read.

Suppose now that the same person is asked to read aloud each word in a sentence, one after the other. Under such conditions the reading will apparently not be automatic because reading of each word is monitored as required by the task. This will result in full representation of each read word up to C3. In particular, the representation at the content level will include a proposition predicating the processed property (meaning) to a specific visual form – the read word.

To sum up, we believe that the term "consciousness" should not be constrained to explicit representations up to C3. It should also refer to explicit representations of aspects of perceived stimuli – the explicit representation of a specific aspect should be equated with being conscious of it.

Nonautomatic processing characterizes the "deliberative" mode of consciousness (Dulany 1991) and results in propositional representations that are monitored by the self. It parallels the notion of an "awareness event" (LaBerge 1997) that results from simultaneous neural activation of two triangular circuits, each connecting three brain sites, one of them common to both circuits and serving as a control area, in the prefrontal cortex. One of these circuits provides a cortical representation of the monitored "object" and the other, the self.

Automatic processing frequently results in explicit representations of only some aspects of the relevant content; it results in explicit representations of properties represented, but not in propositions that predicate these properties to the perceived stimuli. Such representations reflect an "evocative" mode of consciousness that results in conscious representations that are less than propositional (Dulany 1991) and provide only "the sense of" whatever they represent.

NOTE

1. We prefer the term "nonautomatic processing" to "willed action" for reasons to be discussed below.

Implicit knowledge as automatic, latent knowledge

John R. Vokey[a] and Philip A. Higham[b]

[a]*Department of Psychology and Neuroscience, University of Lethbridge, Alberta, Canada T1K 3M4;* [b]*Department of Psychology, University of Northern British Columbia, Prince George, British Columbia, Canada V2N 4Z9.* **vokey@uleth.ca home.uleth.ca/~vokey highamp@unbc.ca quarles.unbc.ca/psyc/higham/**

Abstract: Implicit knowledge is perhaps better understood as latent knowledge so that it is readily apparent that it contrasts with explicit knowledge in terms of the *form* of the knowledge representation, rather than by definition in terms of consciousness or awareness. We argue that as a practical matter any definition of the distinction between implicit and explicit knowledge further involves the notion of control.

One advantage of the natural language meaning of the implicit-explicit distinction as applied to knowledge representations is that it provides a principled explanation for why the implicit is so quiet: It contrasts with the explicit by being in a *form* that cannot be expressed. Thus, rather than "unconsciousness" being a defining (and then yet-to-be-explained) characteristic of implicit knowledge – as in, "implicit knowledge is just like explicit knowledge, except it's quiet" – the "unconsciousness" associated with the implicit is a consequence of this indirect representation (see O'Brien & Opie 1999, and their similar distinction between "vehicle" and "process" theories of consciousness). But perhaps a better term than implicit knowledge for capturing this meaning of indirect representation would be *latent* knowledge, in the natural language sense of "hidden" and "unappreciated." Such knowledge is not merely *implied* (and, thereby, completely without effect until made predicate explicit) by other *explicit* representations, as in Dienes & Perner's (D&P's) bachelor and King of France examples, but rather is indirectly represented because it is distributed over the network of semantic and other (e.g., instance or episodic) data bases. Implicit knowledge as latent knowledge accurately describes the representation resulting from D&P's preferred mechanism of first-order neural networks for the acquisition, retention, and use of implicit knowledge (cf. O'Brien & Opie 1999); it also accords well with the remarkable demonstrations of such knowledge in the large-scale, autoassociative networks of the Latent Semantic Analysis (LSA) models of Landauer and his colleagues (e.g., Landauer & Dumais 1997; also see Laham 1997).

Implicit knowledge as latent knowledge also accounts for the attraction of instance (or exemplar or episodic) models as explanations for implicit learning, as in Brooks' (1978) early memory-for-instances account of Reber's (1967; 1969; 1976) original claims for implicit abstraction of structure in artificial grammar learning. Because the categorical structure is latent in the distribution of instances, learners will behave in a structured manner, even though they are responding only to the memory for individual instances. As with D&P's theory, the knowledge of structure is implicit in instance accounts of implicit learning because it is not directly represented (see also Vokey & Brooks 1992; Whittlesea & Dorken 1993).

D&P's approach emphasises both implicit and explicit knowledge in the positive sense, but in many of the example domains they discuss, especially artificial grammar learning and context-specific item recognition, an important role for latent knowledge may be to support coming to know in the negative sense (i.e., that something is *not*, for example, a member of a category or a previously studied training list): recognising, for example, *only* at a test of face recognition that a test face from a particular minority group (e.g., moustache wearers) could not be a target item because there were no members of that group in the study set, or detecting correctly *only* at test that a test letter string is nongrammatical because it begins with an "X" and none of the grammatical training items did. Because of the possibly infinite number of dimensions of difference between set and non-set members, it would be absurd to suppose that all such dimensions were precomputed and directly or explicitly represented prior to the test. We believe that this test-cued *detection of novelty* plays a major but unappreciated role in many implicit learning tasks that have focused primarily on hits, rather than on the control of false alarms (see Brooks et al. 1997; Higham & Brooks 1997; Higham et al., in press; Vokey & Brooks 1994; Vokey & Read 1995; Wright & Burton 1995).

D&P acknowledge that direct or explicit representation in their theory (i.e., predication explicitness) *by itself* does not *necessitate* conscious access to or awareness of the knowledge so represented (i.e., what they refer to as "attitude explicitness"). As noted, it is also the case in their theory that unconsciousness is a consequence and not a defining characteristic of implicit (latent) knowledge. Thus, as they note in their conclusion (sect. 5), the conscious-unconscious distinction is *at best* only imperfectly correlated with the implicit-explicit distinction. It is surprising, then, that D&P are willing to put so much weight on such evidence as accuracy-confidence correlations, and the "guessing criterion" as diagnostic, especially of implicit knowledge. At best, such evidence implies that the learner has some attitude-explicit knowledge. Such correlations do *not*, however, imply that the knowledge responsible for the residual behaviour is necessarily implicit, any more than they imply that the explicit knowledge is necessarily responsible for the behaviour with which it is correlated; *inter alia* such correlated explicit knowledge could often occur as a consequence of the operation of implicit knowledge, as in our examples of com-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

ing to know what something is not, or that it may be present but not be the functional source, as in Allen and Brooks (1991), for example, in which participants given a simple, explicit rule for categorisation still responded to the specific similarity of the exemplars.

The key concern is that demonstrations such as accuracy-confidence correlations or the "guessing criterion" rely on some form of dissociation logic. As we have seen during the last 30 years of research on implicit learning, critics of implicit learning rarely find such demonstrations convincing. For these reasons we have argued recently (Higham et al., in press; Higham & Vokey 1999) that a more useful definition of the distinction between implicit and explicit knowledge involves the notion of *control,* and a research paradigm that relies on opposition logic based on control (e.g., Jacoby 1991), rather than on dissociation logic based on some measure of explicitness (e.g., verbal report). That is, to be useful, the implicit-explicit distinction must track the automatic-controlled distinction, simply as a practical matter for investigation, if not on logical grounds.

ACKNOWLEDGMENT

This work was supported by operating grants from the Natural Sciences and Engineering Research Council of Canada to each of the authors. Reprint requests should be sent to John R. Vokey.

Questioning explicit properties of implicit individuals in knowledge representation

Carmen E. Westerberg and Chad J. Marsolek

Department of Psychology, University of Minnesota, Minneapolis, MN 55455.
carmen@levels.psych.umn.edu **marso002@gold.tc.umn.edu**

Abstract: Dienes & Perner argue that the explicit representation of an individual to which a property is attributed requires explicit representation of the attributed property. The reasons for this conclusion are similar to the reasons why another of their conclusions may be considered suspect: A property may be explicit without an explicit representation of an individual or the predication of the property to an individual. We question the latter conclusion and draw connections to neurophysiological and cognitive evidence.

Early in their very interesting explication of the explicit-implicit distinction and its application to knowledge representation, Dienes & Perner (D&P) consider that different parts of the content of a propositional attitude (property, individual, predication of property to individual, and temporal context/factuality) may be independently represented explicitly or implicitly. To help systematically organize how knowledge representations may be explicit or implicit, they argue for limits in the possible combinations of explicit and implicit parts of represented content. In particular, an important limit (especially for research on subliminal priming) is that a property may be explicit while the individual and predication remain implicit, but the individual and predication cannot be explicit without an attributed property also being explicit. Alternatively, we suggest that property, individual, and predication cannot vary independently in explicitness-implicitness and that assessments of direct-test performance in subliminal priming research do not actually demand that they do vary independently.

D&P use a pointing-response example to illustrate why explicit representation of an individual to which a property is attributed requires explicit representation of the attributed property. When a person must point to one of two alternatives to indicate which has a particular property (e.g., which has the property of being-a-cat?), pointing to one of the two objects in response to the question necessitates explicit representation of the relevant individual object. In addition, in terms of the knowledge that a person must bring to bear in this task, the attributed property must also be explicit, because the person must explicitly represent the potential attribution of the property to each individual (go into a cat or no-cat state for each individual) to make the correct choice and respond appropriately. Note that the knowledge in the representations that are used reveals what is explicit, rather than what is explicit or implicit in the outward behavioral response. We agree with this suggestion.

However, D&P use another example to illustrate the possibility of an explicit representation of a property without explicit representation of an individual or predication of the property to an individual. When a person must simply name an object in front of him or her (e.g., "cat"), the response necessitates explicit representation of the property (e.g., being-a-cat). But, must the individual or the predication also be explicit? D&P say no, despite the logical possibility that an individual and a predication may be explicit in the internal representation without being explicit in the overt behavioral response. Indeed, we suggest that, as before, what is important is the knowledge in the representations that are used to produce an accurate response (rather than what is explicit in the outward behavioral response). In terms of the knowledge that a person must bring to bear in this task, the individual and the predication of the property to the individual must also be explicit, because the person must explicitly represent the individual to go into a cat or no-cat state as it applies to *that particular individual* (as distinguished from any other – previously or subsequently encountered – individual). Otherwise, it is not clear why the knowledge that a person must bring to bear (for accurate task performance) is used as a critical factor in the pointing example but not in the naming example.

We suspect that explicit representation of an individual (and the relevant predication) is required for explicit representation of a property attributed to it. Thus, the first three elements of content may be all or none, in that all three must be explicit if one of them is. (We should note, however, that we do not take issue with the claim that factuality, attitude, and self may be independently explicit or implicit, compared with the first three elements of content.)

As an aside, neurophysiological evidence may seem to provide, at first glance, an example of a property (alone) being represented explicitly. Certain visual features of the same object (e.g., motion, color, shape, etc.) have been shown to be represented independently in different early streams of primate visual cortex (e.g., DeYoe & Van Essen 1988; Maunsell & Newsome 1987; Zeki 1978). Independent explicit representation of such features (properties) of the same object (individual) may seem to imply the absence of an explicitly represented object to which the features are associated (i.e., an explicit representation of a common object to which the features are associated would seem to violate the independence of the features). However, even in such early visual representations, some explicit information about the object to which a feature is associated (e.g., at least its retinotopic location) must be represented or else there would be no way to eventually bind independent features to the correct, common object (as distinguished from other possible objects). Thus, even in such cases, the explicit representation of the property entails explicit representation of individual-specific information.

Is the possibility of explicit representation of a property without explicit representation of the individual or predication critically important for applying the explicit-implicit distinction to the cognitive literature? D&P suggest that it is important for distinguishing "direct-test" procedures used in subliminal priming research. For example, they describe Marcel's (1983) procedure of asking whether a word (any word) was present or absent as requiring an explicit representation of a predication of the property "word" to an individual (i.e., the stimulus event). We agree. But, according to D&P, that procedure should be differentiated from asking which of four color words was presented (Cheesman & Merikle 1984), for example, because the latter procedure should be understood as requiring only the relevant property (e.g., blue) to be represented explicitly; no explicit predication of the property to an event is required to provide an answer to the question.

However, we would suggest that, in terms of the knowledge that an observer must ring to bear in the latter task, the stimulus event and the predication of the relevant property (e.g., blue) to that event must also be explicit. The person must explicitly represent a particular event as distinguished from any other (e.g., previously or subsequently encountered) event to go into a blue or no-blue state for that particular event; without doing so, the observer would not be able to decide which color word was presented *in that trial.* If both sorts of direct tests require explicit representation of property, individual (event), and predication, what differentiates such tasks so that they could produce different patterns of results? Such tasks may differ in the explicitness with which attitude and self are represented and related with the content, and this may dramatically change the representations recruited for task performance, as D&P convincingly describe for other domains of cognitive processing.

Consciousness and control: The argument from developmental psychology

Philip David Zelazo[a] and Douglas Frye[b]

[a]*Department of Psychology, University of Toronto, Toronto, ON, Canada M5S 3G3;* [b]*Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104.* **zelazo@psych.utoronto.ca doug@psych.nyu.edu psych.utoronto.ca/~zelazo/**

Abstract: Limitations of Dienes & Perner's (D&P's) theory are traced to the assumption that the higher-order thought (HOT) theory of consciousness is true. D&P claim that 18-month-old children are capable of explicitly representing factuality, from which it follows (on D&P's theory) that they are capable of explicitly representing content, attitude, and self. D&P then attempt to explain 3-year-olds' failures on tests of voluntary control such as the dimensional change card sort by suggesting that at this age children cannot represent content and attitude explicitly. We provide a better levels-of-consciousness account for age-related abulic dissociations between knowledge and action.

Many of Dienes & Perner's (D&P's) arguments about the distinction between implicit and explicit knowledge follow in a fairly straightforward fashion from the assumptions stated in their target article, namely, (a) that knowledge should be analyzed according to the representational theory of mind, (b) that some version of the higher-order thought (HOT) theory of consciousness is true, and (c) that action control can be explained by Norman and Shallice's (1980) notion of a supervisory attentional system (SAS). Unfortunately, these assumptions yield incompatible inferences, implying that at least one of them is false.

Assumption (a) provides the basis for D&P's partial hierarchy of explicitness, wherein attitude explicitness entails content explicitness, but not vice versa. However, let us examine what happens when this partial hierarchy is considered in connection with assumptions (b) and (c) and offered as an account of behavioral data.

According to the HOT theory, consciousness of something is tantamount to having a higher-order thought that I am conscious of it. In support of this theory, D&P suggest that "it is inconceivable that one could sincerely claim, 'I am conscious of this banana being yellow' and at the same time deny having any knowledge of whether one sees the banana, or hears about it, or just knows of it, or whether it is oneself who sees it, and so on" (sect. 3.1, para. 2). Now, given that the statement being analyzed ("I am conscious of this banana being yellow") refers explicitly to both self ("I") and attitude ("am conscious of"), it is indeed difficult to imagine such denials, but in this instance, the proposal seems almost tautological.

D&P's subsequent sentence is more substantive, and more to the point: "That is, it is a necessary condition for consciousness of a fact X that I entertain a higher mental state (second-order thought) that represents the first-order mental state with the content X." In this rendition, however, the theory is much less compelling, especially from a developmental perspective. For example, it is hardly inconceivable that 3-year-olds could be conscious of a fact ("There are pencils in the Smarties box") without being conscious of their attitude (belief) or being conscious that they themselves are entertaining the attitude toward the fact (Zelazo 1996; Zelazo & Zelazo 1998). Nonetheless, this is exactly what D&P are claiming; indeed, following Carruthers (1996), they appear to be claiming that it is impossible for an organism even to be capable of conscious predication in the absence of what is essentially a theory of mind.

This analysis fits with D&P's account of the declarative-nondeclarative distinction (sect. 3.3). According to D&P "declarative knowledge represents predication and factuality explicitly, thus qualifying for conscious access" (sect. 3.5). On this account, if we were to find an instance of declarative knowledge, then this knowledge would be conscious and it would be inconceivable that its holder might be unable to access an explicit representation of attitude and self.

At this point, one might begin to wonder whether this version of the HOT theory undermines D&P's partial hierarchy of explicitness (because the hierarchy describes entailments from self to factuality, whereas the HOT theory describes entailments from factuality to self), but the real limitations of D&P's theory are revealed when they turn to developmental data. After establishing that even 18-month-old children are capable of representing knowledge in a fact-explicit fashion (sect. 2.1.1; also n. 6), from which it follows that they are capable of explicitly representing content, attitude, and self, D&P attempt to use the partial hierarchy to account for age-related differences in action control. In the example cited, 3-year-olds perseverate on the dimensional change card sort (DCCS). In the DCCS, children are first told to sort cards according to one dimension (e.g., color, "Put the blue ones here; put the red ones here") and then to switch to another game (e.g,. shape: "Put the flowers here; put the boats there"). Regardless of which dimension is presented first, 3-year-olds typically continue to sort by that dimension despite being able to answer explicit questions about the new rules (e.g., Zelazo et al. 1996).

D&P suggest that 3-year-olds cannot switch because of a failure of the SAS ("Without SAS the once-learned colour sorting rule is dominant and will suppress execution of the new rule"; sect. 3.4, para. 4). The SAS controls schemata via their content and hence requires explicit representation of that content. But D&P also suggest that SAS requires explicit representation of that content. But D&P also suggest that SAS requires attitude-explicit representations: "To avoid confusion, this content must be explicitly marked as being not factual (i.e., explicit representation of factuality), but something that is desired or intended (explicit representation of attitude)" (sect. 3.4, para. 5). D&P then imply that children at this age fail to represent content and attitude explicitly.

Three-year-olds are clearly capable, on the D&P account, of explicitly representing factuality. And in the case of the DCCS, when they are asked directly, children can state the new rules, so their knowledge of these rules is clearly declarative. However, if 3-year-olds represent the rules in a fact-explicit fashion, then according to D&P, they should be able to represent attitude and self explicitly. How, then, can the claim be made that 3-year-olds are restricted to contention scheduling on the basis of the "vehicle" features and so cannot control schemata via their content? When combined with HOT theory, D&P's partial hierarchy of explicitness appears unable to explain the well-documented developmental changes in action control that occur after 18 months of age (see chapters in Zelazo et al. 1999). The implications of D&P's account of explicitness and consciousness contradict the implications of their attempt to invoke Norman and Shallice's (1980) model. Something has to give, and we suggest that the HOT theory is the most troublesome assumption.

The card sorting examples (e.g., Zelazo et al. 1995; 1996), and the presence of age-related abulic dissociations in general, challenge D&P's unitary conception of consciousness. On D&P's ac-

count, there are levels of explicitness, but access consciousness of a representation is something someone either does or does not have [see Block: On a Confusion About a Function of Consciousness" *BBS* 18(2) 1995]. Any representation that is (at least) fact-explicit is conscious, whereas any representation that is fact-implicit is unconscious.

An alternative is to postulate levels of consciousness (Zelazo & Zelazo 1998). Three-year-olds who verbally answer explicit questions about the post-switch rules in the DCCS are clearly in some sense conscious of the rules they describe, but there is another sense in which they fail to reflect further on their conscious state, as shown by their failure to select the rules for action. According to the levels of consciousness (LOC) model, there are four major age-related changes in action control from birth to the end of the preschool years that are explained by increases in self-reflection. Self-reflection occurs via a functional process of recursion whereby the contents of consciousness are fed back into consciousness so that the contents of consciousness at one level (or moment) become available to consciousness at a higher level. With each new level of consciousness, children are able to exercise a new degree of control over their behavior because they can formulate rule-governed actions of greater complexity and can maintain those rules in working memory, as captured by the Cognitive Complexity and Control theory (Frye et al. 1995; Zelazo & Frye 1997). This approach allows us to account for age-related changes in action control in tasks such as the DCCS and to trace the development of adult-like consciousness using very few theoretical tools.

Authors' Response

Deconstructing RTK: How to explicate a theory of implicit knowledge

Josef Perner[a] and Zoltan Dienes[b]

[a]*Institut für Psychologie, Universität Salzburg, A-5020 Salzburg, Austria;* [b]*Experimental Psychology, University of Sussex, Brighton, Sussex BN1 9QG, England.* **josef.perner@sbg.ac.at dienes@epunix.susx.ac.uk www.sbg.ac.at/psy/people/perner__e.htm www.bids.susx.ac.uk/faculty/ep/dienes.htm**

Abstract: In this response, we start from first principles, building up our theory to show more precisely what assumptions we do and do not make about the representational nature of implicit and explicit knowledge (in contrast to the target article, where we started our exposition with a description of a fully fledged representational theory of knowledge (RTK). Along the way, we indicate how our analysis does not rely on linguistic representations but it implies that implicit knowledge is causally efficacious; we discuss the relationship between property structure implicitness and conceptual and nonconceptual content; then we consider the factual, fictional, and functional uses of representations and how we go from there to consciousness. Having shown how the basic theory deals with foundational criticisms, we indicate how the theory can elucidate issues that commentators raised in the particular application areas of explicitation, voluntary control, visual perception, memory, development (with discussion on infancy, theory of mind [TOM] and executive control, gestures), and finally models of learning.

R1. Deconstructing RTK (representational theory of knowledge)

Several commentators have criticised us on points that seem to be consequences of our adopting RTK as a framework for our exposition. In fact, we had not explicitly used RTK (or RTM; Representational Theory of Mind) in our original draft. We introduced it in a revision with the aim of providing readers with a familiar framework detailing the elements of propositional attitudes, without wanting to buy into the usual interpretation of being like language (Fodor 1975: "A language of thought"). In other words, our strategy (as it finally appeared) was top down, to start with the most explicit and elaborate human understanding of knowledge and then decompose it into its elements. Since the starting point is highly permeated with language, this created the wrong impression of what we are trying to achieve. So, we take to heart **Gall**'s admonition that we rely too heavily and too early on RTK and **Carlson**'s advice to invert our focus.

So, we now try to trace our enterprise in the opposite direction, from the bottom up. This may help allay some of the fears about the core assumptions underlying our analysis. To overview the issues: Whether we start top down or bottom up, we do presume that having knowledge or holding a belief involves explicit representation, and to that degree we hold a representational theory of knowledge. A fully fledged RTK holds that one can know a proposition p only if p is itself explicitly represented. Thus, for fully explicit knowledge, we hold that RTK is strictly true. For implicit knowledge, we also hold there must be explicit tokening of some representation. But in contrast to RTK, we allow implicit knowledge that does not consist in a representation tokening the full proposition p. It is only to that degree that our framework is not a RTK. Our commentators must bear in mind that subscribing to these assumptions does not entail subscribing to all other assumptions of a Fodorian world view, for example, the assumptions of RTK to which we do subscribe can be (perhaps should be) held even by a rabid connectionist, a point to which we return in section R7 below.

R1.1. Overly linguistic

Several commentators complained that our analysis of implicit-explicit knowledge is too linguistic (**Carstairs-McCarthy, Pietroski & Dwyer, Jiménez & Cleeremans**) and anthropomorphic (**Mercado & Murray**) because we are relying heavily on the representational theory of mind (RTM) or knowledge (RTK). It is true that we start from an analysis of ordinary language expressions about the mind (that's what philosophers of mind are mainly engaged in). But this starting point is hard to avoid. Even behaviourists usually rely on anthropomorphic descriptions of what the animal is doing: pressing a lever, jumping through a hoop. Of course, as research progresses it moves away from that starting point and develops better analyses for the specific matter of investigation. On occasion, however, it is important to remind oneself of the starting point, because dedicated research often forgets some useful distinctions. For instance, memory research for many years had lost the distinctions that are re-evoked in the implicit-explicit and in the semantic-episodic distinction (Tulving 1985) and that were originally cited by the old masters, for example, Ebbinghaus (1885) and James (1890).

Evidently, these distinctions have been made primarily on the basis of our linguistic distinctions and our phenomenology. There is no reason, however, why these distinctions could not be separated from their linguistic and in-

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

trospective origins in order to investigate the presence of these processes in non-linguistic animals. One example of such an enterprise is the work of Dickinson (e.g., Heyes & Dickinson 1993), who asked whether or not rats represent propositional attitudes, like their goals and intentions. One of our objectives is to analyse why in the human case of language use, consciousness, voluntary control, directness of tests, and so on (a point appreciated by **Evans & Over**) tend to go together; this is so that we can design experiments that do not rely on linguistic competence.

R1.2. Causally inert

At the bottom level we are concerned with representations. As a quick definition: representations are states (typically internal) of an organism that are about something (typically the organism's environment). They get their "aboutness" from the fact that they causally govern the organism's interaction with its environment by mapping the relevant distinctions in the environment. They can only map the environment non-accidentally if there is a causal process from environment to representation (e.g., perception). Environmental differences that are reflected (encoded) in the representation are represented explicitly. The interesting thing here is that even if my representational capacities only allow me to make a distinction between *lion* and (domestic) *cat* which then controls relevant behavior, if the cat in front of me makes my mind go into its cat state, then that state represents *implicitly that there is a cat* and not just *cat-ness.*

Even though it is as implicit as they come, this representation is *not* causally inert, as **Jiménez & Cleeremans, Carlson,** and **Vokey & Higham** (latent knowledge – completely without effect) suggest. What one could say is that implicit knowledge has fewer causal effects than more explicit knowledge, since the latter allows more internal distinctions, which can lead to a greater variety of causal effects. But implicit knowledge is not causally inert! In defence of these claims of causal inertness, one could surmise that these commentators interpreted "implicit knowledge" as referring only to those aspects of implicit knowledge that remains implicit. Since these aspects are not reflected in internal differences, they cannot have any causal consequences for behaviour – such appears to be the reasoning that leads these commentators to claim causal inertness. However, even this is not quite right. The reason the implicit aspects are "represented" at all is that they are involved in the causation of the internal representation: if it weren't for the fact that it was the particular cat that was responsible for my mental "cat" token, then the fact that it was that particular individual, which happens to be a cat, would not be implicit in my explicit "cat." Moreover, the causal role of the implicitly represented individual is also important for the appropriateness of my behavioural effects of the explicit parts, for example, saying "cat" and smiling, as opposed to saying "lion" and running away in fright. If there was no particular individual or if the situation were not real, then my behaviour would be inappropriate. What the implicit-explicit distinction captures is where the causal effects are located: in the environmental setting (implicit) or in the internal distinctions (explicit). It thus captures an important aspect of the substance matter, that is, how different kinds of knowledge can be used, and is not just a theory of how scientists use the terms, as **Taatgen** suggests.

R1.2.1. The implicit piggybacks on the explicit. At this point one may also wonder whether **O'Brien & Opie**'s "implicit knowledge contents piggyback on the explicit" is an accurate characterisation of our position. One interpretation of this is that implicit aspects depend counterfactually on explicit aspects. True, if there were no explicit representation of lion versus cat then there would be no aspects implicit in anything. However, if the source of the implicit aspects – that is, the fact that the particular individual is the cause of the explicit distinctions – did not exist then there would be no explicit distinction. So the explicit is also piggybacking on the implicit.

R1.2.2. Explicit individuals with implicit properties. The causal role that properties and individuals play in knowledge formation provides a good context for addressing the question of whether there is such an implicit-explicit hierarchy that properties can be explicit, with individuals as the carriers of that property remaining implicit, but not the other way around. As **Barber** correctly observes, the main purpose of our analysis is to lay open the possible elements according to which knowledge can be implicit or explicit. Nevertheless, we also had the intuition that not all combinations are possible and tried to formulate a partial hierarchy. Barber agrees with our intuition that there is some asymmetry but challenges our specific proposal with a counterexample. In his variant naming game, the player is confronted with several individuals of whom one is being highlighted at each turn. The player just identifies the particular individual explicitly (mentally as well as verbally), but makes no internal distinction concerning the property of *being highlighted.* The player relies implicitly on the fact that his identification is being taken to refer to whatever is being highlighted.

We agree that this is an intriguing counterexample and our answer is not one of the two nonviable options anticipated by **Barber.** Rather we wish to point out that the counterexample only seems to work with specific properties such as being highlighted, which is not primarily a property of an object but a property that describes the interaction between the object and the players of the naming game. The lesson we take from this observation is that whether or not something can be left implicit depends on the causal relationship between these aspects and the observer (game player). Because the highlighting causes the player to attend to the particular individual, the property of being highlighted can be left implicit. In general, however, there remains an asymmetry between individuals and properties: it is necessary that some property be represented explicitly, namely, the one that individuates the object in the observer's mind, before any individual can be identified.

Westerberg & Marsolek deny that either the individual or the property can remain implicit, because in the naming game the player must represent that "cat" applies to *that particular individual,* and in the subliminal Stroop experiments one must represent which colour word was presented *in that trial.* Our point, however, is that if one does not predicate the perceived properties of any particular event or individual but simply answers with whatever colour first comes to mind, one can do better than chance because the most recently presented colour word makes it more likely that it comes to mind first. The subject then makes an inference, attributing the colour to a particular

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

trial, but this inference occurs some time after the moment of perception itself.

The physiological evidence mentioned by **Westerberg & Marsolek** is that visual properties are encoded separately and later bound together (predicated of a single individual); this illustrates the possibility that on occasion the property information alone might make it into higher brain regions, without the binding information. This could still have some behavioural effect, whereas if the property information is lost and only the binding marker survives then it is hard to imagine what behavioural effect this could have. In early vision, location is initially coded property-structure implicitly in spatiotopic feature maps where there is no single representation for a particular location, but many location-feature representations. Hence the individual is not coded explicitly, but binding to an individual object is still possible at a later stage of processing, as the result reported by **Bridgeman** suggests.

R1.3. Property-structure, predication and non-conceptual content (NCC)

How does nonconceptual content (NCC) fit into our framework, **Brinck** asks. NCC bears an interesting relation to property-structure implicitness. Chrisley (1996) defined NCC as content that is not entirely composed of constituents that meet the generality constraint (the constraint that constituents can freely recombine with each other). A representation that carries NCC with constituent structure would thus be property-structure implicit. Suppose the nonconceptual content in question was *green and small,* which is NCC if the constituents do not satisfy the generality constraint. So *green and small* is not represented by an all-purpose *green* token concatenated with an all-purpose *small* token. Thus, the structure of being green and being small is not made explicit by the representation of *green and small*; it is property-structure implicit. The representational content also meets the definition of NCC given by Brinck because the holder of this content can have it without having the concepts *green, small,* and so on that we use to describe the content.

On Cussins' (1929) view, NCC cannot be predicated of an external (conceptually identified) object (since NCC does not necessarily respect the boundaries of such objects). Consequently, NCC cannot have a truth value, because only expressions that predicate properties of individuals have a truth value (Evans 1975) in the classical sense of being able to derive contradictions. This view conforms with **Brinck**'s characterisation of NCC as having correctness conditions without being able to have a truth value assigned. When this is combined with the claim that NCC is accessible to consciousness and volitional control it poses a problem for our claim that explicit predication is a prerequisite for consciousness and volitional control.

Several theorists take a different view, however. Chrisley (1996), Peacocke (1993), and Bermudez (1995) do regard NCC as propositional and capable of having a truth value (there is a way of predicating that allows this). Thus, on these views, NCC poses no problem for our framework: It may be represented maximally implicitly as a property, or fully explicitly, as a representation of knowing an individual has a certain property (conscious but not verbalisable because the property cannot be conceptualised).

NCC interpreted as structure-implicit representations also makes it clear that our immediate action regulation is based on NCC: Our interaction with the world involves representations that structure-implicitly represent a mix of object properties and features indicating how to act on these objects, because this is the most efficient way of effecting action (e.g., common coding of perception and action; Prinz 1990). Normal action is therefore difficult to verbalise, as NCC cannot be dissected with our concepts. However, under the assumption that NCC is predictable, this, as **Brinck** observes, allows for the intentional and willful improvement of craftsmanship through perceptual monitoring in the absence of verbal reflection.

Why should verbal reflection come into the picture? We suspect it is because the mention of predication and propositional conjures up images of "language-like representation" as in the analogue versus propositional representations dispute (Kosslyn 1975; Pylyshyn 1973). There is of course some link between the propositional and the linguistic. Linguistic expressions are characterised by a high degree of articulation of their meaningful parts, that is, basic units of meaning (words) are linked by precise rules of concatenation into larger meaningful units (sentences). Images as prototypical analogue representations are meaningful without having any clearly separable parts. For an explicit representation of predication a minimal degree of articulation is needed to link the predicate to its subject. No further degree of articulation is needed for the predicate, however. It could be an image. In any case, in this view predication is something very fundamental and not just a feature of language, as **Carstairs-McCarthy** puts forward, and it is something of which animals must be capable if they engage in variable binding regardless of their linguistic abilities.

For example, summaries of various features of NCC (Brinck 1997; Peacocke 1993) list the finer grain of visual images as one feature of NCC. Like the detailed imagistic schema of faces by which we are able to recognise so many different people, the content of images consists of properties that can be predicated to objects or events in the world. And because their content can be predicated, this predication and its factuality and eventually our knowledge thereof can be made explicit and consciously experienced, even if they cannot be completely described.

R1.4. Concepts and property structure

In our view, one has a concept of a property only if one has the internal distinction whose function it is to indicate that property. This is defined purely by its semantic/symbolic relation to the world. Hence it is a distinction which is predicable of the particulars in the world that carry the property. However – and Fodor could not object – these conceptual distinctions can only fulfill their semantic function if they are embedded in processes among which other distinctions are made, many of them being non-conceptual and property-structure implicit in a way that cannot be explicated. Since it is implicit, it cannot be coherently addressed for different purposes, which may explain why people give idiosyncratic responses when asked about it and produce incompatible results for different tasks such as rank ordering definitional properties as opposed to rank ordering category instances by typicality, as observed by **Hampton.** Conceptually defined criterial properties may play little role in typicality judgements.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

In this context **Hampton** raises the difficult question of what properties are structure implicit in other properties. He suggests that all contingent implications, such as *being composed of cells containing DNA,* are property-structure implicit in *bachelor.* This seems to go too far, violating the linguistic intuition of what is conveyed implicitly when one says "He is a bachelor." This does not implicitly convey that he is made up of DNA-carrying cells. To avoid this consequence, we formulated our criterion in terms of meaning. It is not just any supporting fact that makes for implicitness; only the ones "that are necessary for the explicit part to have the meaning it has" (sect. 1, penultimate para.).[1]

The distinction between analytic and synthetic truths has been criticized by Quine (1951). However, the intuition behind the distinction does not go readily away. Keil (1989) has made good use of a distinction between definitional and characteristic features in children's acquisition of concepts. This distinction underlies the strong intuition that sensitivity to some features but not others is essential for a proper understanding of a concept. As far as we can tell, the question of how to make this distinction is unanswered. We can rely only on our natural linguistic intuition.

One interesting question concerns the role the distinction between defining and characteristic features may play in an "externalist" view of concepts like Fodor's, in which the concept is purely determined by its semantic relations to a property. One possibility (Keil 1998; Perner 1998) is that defining properties need to be internally distinguished so that the target distinction can serve its representational function. In that case the concept *bachelor* would necessitate conceptual or nonconceptual sensitivity to *maleness* and being *unmarried* but no such sensitivity for detecting the presence of DNA, cells, and so forth. Moreover, despite the required sensitivity to maleness and being married, no definitions in terms of the corresponding concepts need to be formed.

Logical implications are likewise unnecessary for meaning. A mathematician who knows Peano's axioms does thereby not implicitly know all the truths entailed by them. So, in our definition of property-structure implicitness in the case of Plato's Meno, where the young boy is led by his teacher to draw out the implications of what he already knows, is (in agreement with **Homer & Ramsay**) not a case of making property structure explicit. Contrary to Homer & Ramsay, conscious reflection is sometimes not enough to make property structure explicit, as we discussed in the case of NCC. By way of an empirical example (of, as it were, nonconceptual content [NCC] relative to a specific domain and task), Roberts and McCleod (1995) found that people trained under full attention to recognise exemplars of the category, for example, "triangle and red" were equally good with a monochrome display which only showed shapes without colours in reporting triangles as *possible* instances of the category, but they were rather poor at recognising the triangle as a possible instance after learning with diverted attention.

Contrary to **Overskeid**'s claim that representing a compound property ipso facto explicitly represents its components, the Roberts and McCleod paper shows that property structure implicitness is not only logically possible but empirically observable. (This can be achieved by representing the components in a context sensitive way; i.e., their only function is to indicate the component when the other components are present; thus, each component is not explicitly represented in itself.)

R1.5. Factual, fictional, and functional use of representations

At bottom, a strict separation of representation and functional use is not possible because (by our provisional definition) representations do not just map the environment but also govern the interaction with that environment. A relative separation of representations from their use emerges in more complex systems as the articulation of components increases. Imagine a connectionist robot that can learn to negotiate an environment to get to a particular goal. The representation of the goal may be enmeshed with the representation of the given environment because when the goal changes the robot has to relearn a great deal about the layout of the environment. In this case, there is no systematic internal distinction between the two basic functional uses: beliefs and desires. This distinction is property-structure implicit in a representation in which information about the environment and about where the robot wants to be are inextricably enmeshed. For a system that can flexibly combine knowledge of the environment with its goal, this separation needs to be made by some internal distinction. Goal devaluation studies seem to show that rats can make this distinction (Heyes & Dickinson 1993). Propositional attitudes, even though they come from a "linguistic" analysis, can be studied in nonlinguistic animals, *pace* **Mercado & Murray** and **Carstairs-McCarthy.**

The necessary internal distinction can be implemented in many different ways, in the philosopher's favourite metaphor of a belief and desire box, or as functional markers on individual representations. Its prime purpose is to ensure the proper use of the representations. However, since by doing so it also classified the marked representations as representing the organism's environment or goal, these functional markers (for beliefs and desires) also qualify as procedural knowledge of the distinction between facts and goals. They are not declarative knowledge. To become declarative the markers themselves have to come within the scope of the belief marker. Only then does one know (believe) that something is a fact or a goal.

In the target article we were concerned less about the belief-desire distinction than about the further distinction between factuality and fiction. The problem can easily be seen. With solely a belief-desire distinction I can only know (believe) or want something. I cannot just think of something. There is one special possibility, however: unpredicated properties. Because they are unpredicated they do not describe a fact, hence they remain nonfactual (but they are not exactly fiction, either). The question of how to introduce the factual-fictional distinction properly has recently been discussed by Nichols and Stich (1998, unpublished manuscript) and by Currie and Ravenscroft (in press b, Ch. 5). Nichols and Stich suggest introducing a third box, namely, a possible-world (PW) box (or type of functional marker – perhaps the omission of one of the other two). We thereby gain a functional distinction between factuality and fiction.

We agree that such a functional distinction is at the heart of the factual-fictional distinction (and all hypothetical reasoning, as **Evans & Over** point out), just as it is for the fact-goal distinction. Hence we agree with **Currie** that we can

not capture factual or fictional status purely within the content; it can only be captured by the functional distinction. However, as our bottom-up analysis – pursued here – shows, the functional distinction implies a representational distinction; the functional marker makes the distinction explicit, although only as a property without explicit predication. This amounts to predication-implicit procedural knowledge of the distinction. In other words, making factuality explicit means introducing a functional (representational) distinction that has the appropriate effects. **Currie**'s question only arose because ours was a top-down analysis with RTK as a starting point.

The bottom-up analysis raises another interesting question not apparent in our original treatment. Is a purely functional distinction providing procedural knowledge of the factual-functional distinction sufficient? This question was recently put into focus by Nichols and Stich (unpublished) in a discussion of pretend play. When pretending that this (banana) is a telephone, infants simply switch to a different functional mode concerning the representation, "the banana is a telephone," without knowing that they are pretending (since the functional use is not registered within the belief box).

Although this is perfectly possible, the intuition among developmental psychologists (e.g., Leslie 1987; Piaget 1945) is that pretence emerges with the knowledge that one is pretending (in some minimal sense). Piaget spoke of the infant's "knowing smile" as an indicator of this reflective awareness. Moreover, Nichols and Stich's suggestion puts hypothetical reasoning, including pretence, on a par with the belief-desire distinction. It follows that our pretence should be able to remain as unconscious as our desires. However, although in our many automatic actions we are often unaware of our reasons for doing what we are doing, the same cannot be said for pretence.

Like the developmental intuition, our phenomenal self insight suggests that pretence (and hypothetical reasoning, etc.) does not occur unconsciously. The fact that we are pretending is always within our belief box. Perner (1991, Ch. 2) – following Leslie's (1987) analysis of pretence – suggested that the real-hypothetical (i.e., factual-fictional) distinction is based on meta-representational context markers which serve a functional and representational role (see also Sperber 1997 for a similar suggestion). In our current terminology we can say that the factual-fictional distinctional does not emerge first as procedural knowledge, but comes directly as declarative knowledge.

Pursuing the option that the factual-fictional distinction consists of a functional distinction within the belief box (or within the scope of the fact marker that distinguishes facts from goals) we can answer another critic. An organism that only distinguishes functionally between facts and goals (belief and desire box) cannot represent the fictional vis-à-vis the factual. For such an organism, **Nichols & Uller** propose a standard rule – if p is believed then one can add "It is a fact that p" – is perfectly possible but useless, since every occurrence of p in the belief box has the function of being taken as a fact. The dorsal action system may be of this kind, provided it processes propositions at all. The rule, however, fails when it becomes relevant in an organism (or our ventral visual information processing path) that can distinguish between fact and fiction with appropriate functional markers. If that organism encountered some proposition p without a marker, then the rule would be dangerous to apply. The claim is that such propositions can float around in our head (belief box). They constitute implicit knowledge of the fact that p, because they have been properly caused by perceiving p, but their factuality has not been explicitly marked. So they remain factuality-implicit knowledge.

The dependence of the factuality-fiction distinction on markers in the belief box makes the distinction akin to that between temporal contexts: knowing what happens now and knowing what happened earlier. There is some evidence that the ability to distinguish fact from fiction and the ability to represent temporal contexts goes hand in hand. As **Boucher** points out, explicit representation of time is part of a cluster of abilities that is controversial concerning (1) whether or not animals possess these abilities, and (2) the difficulty autistic children have with this cluster of abilities. There is also evidence from the study of normal development that the ability to pretend emerges at the time children can represent earlier states of affairs to understand invisible displacement of objects (Perner 1991). Our disagreement with Boucher is that it is not clear to us whether explicit representation of time is the driving force behind these new abilities rather than the more general ability to differentiate contexts within the belief box. We also have difficulty seeing how the development of time-keeping mechanisms provides an explicit representation of temporal contexts.

R1.6. From predication and factuality to consciousness

As we introduced RTK into the revision of the target article we also edited the issue of how to define implicit and explicit knowledge. The relevant passage in the original target article clarified how the implicit-explicit distinction – defined for linguistic expressions and representations – applies to knowledge, a transition that **Gall** found wanting.

> Knowledge of a fact or an aspect of a fact is *explicit,* if that fact or aspect is represented by an internal state whose function it is to covary with it. Other supporting facts or aspects of facts that are not explicitly known but must hold for the explicitly known fact to be known, are *implicitly known.* (Original draft of target article.)

We refined the notion of knowledge by specifying 4 conditions (sect. 2.1.2). **Smith** objects to these conditions because in his view they conflate two standard accounts of knowledge. Indeed we did not spell out the relationship between representation and content in any detail; we only indicated it. For example, in the formulation of "(i) R is accurate (true)," the parenthesis is to indicate that "the proposition represented by R is true," just as Smith suggests. Our four conditions specify primarily the causal account and the information in parentheses or in subordinate clauses indicates how the particular causal condition relates to logical/foundational factors. We can not see why this is objectionable. Our approach, far from conflating two theories of knowledge, appropriately allows a person to believe either theory of knowledge. In any case, however we specify conditions of knowledge, it is difficult to see how that would invalidate what we have to say about the implicit-explicit distinction.

We also argued that making the attitude of knowing explicit requires that the content – in particular, factuality and predication – be made explicit (sect. 2.1.3). Several commentators suggested counterexamples to this claim. **Bibby**

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

& Underwood point out that one can represent "I know that X has the property Y but I don't know what Y is," or, more concretely, one can represent, "I know that this person has a name, but I don't know what it is." The commentators then suggest that this would violate the proposed hierarchy because explicit representation of attitude (I know) is possible without explicit representation of what Y is. A violation of the proposed hierarchy would only be a risk if explicit representation of attitude ("I know that this person has a name") constitutes knowledge of the person's name without making the name, or the fact that the person has this name explicit. This is like in the naming game where "cat" constitutes knowledge of the fact that the animal in front of me is a cat, without making this predication explicit. The proposed example, however, simply does not constitute knowledge of this kind.

A more plausible case is the "feeling of knowing" or "tip of the tongue" phenomenon: "I know this person's name, but it won't come off my tongue right now." Now this complicates the picture, for this phenomenon introduces a distinction between what one knows *long term* and what one knows as *instantly* available. If we construe the "knowing" as long term, then there is no threat to our hierarchy, for somewhere in the mind there is an explicit representation, "This person's name is Susan." Construed in terms of immediate availability, the explicit representation "I know that person's name" does not constitute immediate knowledge of that person's name; in a similar way, the representation, "He knows that person's name" does not constitute knowledge of that person's name. Hence, there is no violation of the proposed hierarchy.

Nichols & Uller mount a different attack on the proposed hierarchy by showing that animals who presumably lack the capacity for explicit factuality nevertheless represent their mental state of perceiving an event explicitly, as shown in the experiments by Cowey and Stoerig (1995) on monkeys with unilateral lesions of the visual cortex. There are two ways in which our analysis can be applied to these findings.

1. We can go along with a rich interpretation that monkeys represent events as visual (explicit attitude) but deny the assumption that monkeys are incapable of explicitly representing factuality. What is the evidence that they cannot? One kind of evidence would be the lack of pretend play. Even anecdotal evidence for such an ability in apes is scarce (Byrne 1995), let alone reliable experimental evidence. This does not mean, however, that primates are incapable of representing factuality. Like children with autism (Lewis & Boucher 1988), whose capacity to represent factuality one does not want to deny altogether, apes may see no point in pretence.

2. We can accept that primates are not able to represent factuality but deny that the study by Cowey and Stoerig establishes that monkeys represent their attitude towards visual events. As **Nichols & Uller**'s careful formulation of "lights" and "nonlights" already suggests, it could be that in the second part of the experiment monkeys press the "light" button not because they represent that they have *seen* something, but because of the presence of some event with a certain property, for example, something shiny (which we call light). In their blind field they do not perceive lights as shiny (hence they do not press the "light" button), but they do perceive other properties, such as its position and that it is a button to be pressed. Or the dorsal system deals with predication-implicit representations, and thus has not predicated all the features of an object or event that are distinguished; without the ventral system, the monkey perceives neither objects nor events as coherent entities.

None of the solutions commits us to assuming that the visual system is drastically dissimilar between humans and monkeys, except for the differences inherent in the assumptions. In particular, if we assume (as **Nichols & Uller** seem to do) that monkeys are incapable of explicit factuality, then unlike in humans, their ventral path evidently does not serve this purpose. Moreover, if explicit factuality is required for consciousness, then the monkey's ventral path must differ from the human's in that it does not provide a conscious experience of the perceived events. In other respects, however, the functions of the ventral and dorsal pathways may be the same in these species.

Nichols & Uller present a second counterargument, along similar lines, pertaining to declarative memory in humans and monkeys. In animals, the hippocampus seems to be responsible for creating memory of conjunctions of features that can be dealt with flexibly (Squire 1992), but there is no evidence that it creates memories that declare something to be so. On the other hand, in people, the memories formed by the hippocampus are genuinely declarative. A memory system built for dealing with one-off conjunctions in a flexible way was perhaps the most suitable starting place for evolution to mould a genuinely declarative memory system in Homo Sapiens.

Mercado & Murray also wonder to what extent dolphins have propositional knowledge; and, even if they do not, whether dolphins are able to represent their attitude of uncertainty explicitly. Mercado & Murray point out that dolphins choose to escape from conditions of uncertainty. Does this indicate representation of a propositional attitude? Maybe not: The dolphin may escape uncertainty not because it has represented itself as being uncertain (i.e., not because it has attitude-explicit knowledge), but because of other effects that the uncertainty has on the dolphin; for example, aversive physiological effects. A better way of getting at attitude explicitness in animals may be to train them to respond with different levers when events happen with different long-term probabilities and then see whether the animal can transfer those responses to assessing singular events. The responses then form confidence ratings for the event happening. A similar methodology is used with children to test their awareness of uncertainty in the context of implicit knowledge (**Ruffman; Goldin-Meadow & Alibali**).

Zelazo & Frye charge that we undermine our proposed hierarchy of explicitness by considering the possibility that explicit factuality might be *sufficient* for explicit representation of attitude (hence consciousness) because one can infer from something being a fact that one knows it to be a fact (by applying an ascent routine: Gordon 1995). This ability to infer, however, is quite different from the hierarchy of explicitness. We are not claiming that one could not explicitly represent "Fb is a fact" without also making "I know . . ." explicit. Because we are not claiming this, we do not undermine the hierarchy. Our claim is only that although there is the possibility of representing factuality explicitly and leaving the attitude of knowing implicit, it may be difficult to detect actual instances of this because people, when questioned can infer and then explicitly represent their attitude of knowledge for fact-explicit knowledge.

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

To incorporate consciousness into our picture we rely on the higher-order thought (HOT) theory of consciousness. We are committed to this theory because we cannot conceive of being conscious of some fact without the ability (requiring a higher order mental state) to specify the first-order mental state by which we behold this fact. **Zelazo & Frye** even find this observation "almost tautological." Yet, when we capture the generality of this observation in the principle that it is a necessary condition for consciousness of a fact X to have a second order thought about the first order mental state with the content X, they object that it is not very compelling.

O'Brien & Opie are right, HOT (higher-order of thought theory) is a controversial theory in the philosophy of consciousness. To a large degree, this controversy is a result of relying exclusively on phenomenal intuition. This is undoubtedly an excellent starting point but at some point of refinement, introspective intuitions become inconclusive because people tend to have different intuitions. (This was our reaction to Block's [1995] examples of the separability of access and phenomenal consciousness.) Further progress will require a predictive theory that goes beyond direct intuitions. Implicit and explicit knowledge is such a domain because it ties consciousness (as a particularly strong form of explicitness) to other distinctions such as directness of test ability, voluntary control, and hypothetical reasoning. The main purpose of our contribution is to explain how these various factors relate to one another and form clusters – a point particularly appreciated by **Evans & Over.** HOT theory does not provide us with these connections, HOT theory only ties consciousness to explicitness of attitude and does not bear the real burden in our project as **O'Brien & Opie** claim. In fact, our project may help vindicate HOT theory if, with its help, we can make the correct empirical predictions.

Even without a HOT theory of consciousness, we can bring order to the empirical facts including those of artificial grammar learning. If one wishes to treat guesses and forced choice responses as evidence of pre-existing conscious explicit knowledge, knowledge that does not require HOTs (even though conscious) that is fine; it is almost just a terminological issue (as **Zelazo & Frye**'s "levels of consciousness" suggest). But the facts are that in artificial grammar learning and other paradigms, subjects acquire knowledge about which they lack HOTs (is attitude implicit), and this is exactly what our framework clarifies. It also seems quite natural to call such knowledge unconscious, even though it does affect conscious experience: Task demands lead it to affect conscious experience downstream of processing (e.g., as preferences, but not as experienced knowledge).

With respect to preferences, **Bornstein** asks whether our framework deals only with propositional knowledge rather than implicit affects or motivational states. Experience can causally influence our affective and motivational states, leading to knowledge of the states' existence (I know that I have the property of being in state X, where X is liking an object), without there being knowledge of having experienced the object before (i.e., there is no retrieval of the representation formed during the perception of the object: "I see that this object has this structure"). Only a representation of the structure need be formed; it need not be predicated of a particular object at a particular time. For affective states to be altered by visual experience, there is no need to represent having seen the object, or to represent that the affective experience is linked to having seen the object. Thus, implicit representations can (through task demands) ultimately lead to some sort of conscious experience that has its own attitude-explicit representation associated with it even though the conscious experience is not one of knowing.

Tzelgov et al. suggest that even predication-implicit knowledge can be conscious. In support of their claim they cite the results of Tzelgov et al. (1997), who presented subjects with one of eight words for different durations. Subjects showed a Stroop effect only when they could report which of the eight words was at above chance levels. This result is entirely consistent with Cheesman and Merikle (1984) as discussed in the target article – the word report task is a test of objective threshold and can be performed with a predication-implicit representation. Tzelgov et al. claim that the subject is conscious of the word because the subject reported it. The subject *is* conscious of the word or is at least led to be conscious of it by the task demands, and in so doing forms a relevant HOT ("I guess the word could have been blue"). This occurs, however, as an act of inference some time after the moment of perception, so the perception itself is unconscious (no HOT involving the attitude of seeing is caused by the perceptual act directly). This process corresponds to Dulany's (1991; 1997) evocative mode of consciousness. The person becomes conscious of something but not of perception per se. Furthermore, the conscious experience arises because of the inferences caused by task demands, not directly because of the predication-implicit representation formed by perception. The experiment by D. G. reported by Tzelgov et al. (using the presence of semantically similar false alarms on a recognition test to indicate the lack of predication during perception) says more about memory than perception. To see this, consider their last thought experiment in which they argue that a person forms a fully explicit representation. If, after a delay, synonyms were given as false alarms more often than control words, would Tzelgov et al. argue that perception was actually predication implicit after all?

R2. Rendering knowledge explicit

Several commentators highlighted the question of how implicit knowledge can be made explicit (Karmiloff-Smith 1992) as an important topic to address. **Georgieff & Rossetti** ask whether all implicit knowledge can be made explicit, and if not why not. Knowledge in the dorsal stream apparently cannot be. How does this figure in our scheme? As noted by Georgieff & Rossetti, the dorsal and ventral systems differ in more ways than just the implicit/explicit status of the knowledge. It is quite possible that when there are independent systems like this one, implicit knowledge in one system has no means of being made explicit. Perhaps a crucial feature is time, as recognised by **Carlson** as well as **O'Brien & Opie.** If a representation is used by the visual system for a short period of time, this may be long enough to exert influence on subsequent processing, but too short to allow predication, factuality, and so on, to be represented. Other representations whose property structure cannot be made explicit are those that carry NCC (nonconceptual content) with constituent structure (thereby providing a counterargument to **Homer & Ramsay**'s claim

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

that knowledge which is property-structure implicit can be made explicit by conscious reflection).

Weights in a connectionist network provide an example of this. Also, within the standard processing assumptions of connectionist networks, there is no easy way for the representational content of weights to be represented as factual or not (Dienes & Perner 1996). Apparent cases of procedural knowledge embedded in weights being made explicit by reflexive abstraction (**Homer & Ramsay**) could just be additional representations formed by hypothesis testing rather than the content of implicit representations being made *directly* explicit ("directly" in the sense that the mechanism predicating etc. the property embedded in the weights is so reliable in detecting the right property that it does not need to *test* whether the property is right). On the other hand, **Carlson** provides an informative analysis of four different ways in which other implicit knowledge may be made explicit.

R3. Voluntary control

A few commentators noticed that involuntary processes are often associated with conscious experience. **Kinoshita,** for example, suggests that involuntary recollection poses a problem for our framework because the recollective experience implies the memory is fully attitude-explicit; but, according to us, volition is also associated with full attitude-explicitness. The answer is that according to our framework, a fully explicit representation is necessary for volitional retrieval but such a representation does not necessitate volitional retrieval. Thus, retrieval volition requires consciousness, but conscious awareness of an item having been on the list does not require retrieval volition. Similarly, **Tzelgov et al.** point out that in a Stroop experiment, reading the word for meaning is involuntary, but the meaning still becomes conscious. Again, on our scheme there is no reason why involuntary processes should not be conscious. We do claim that automatic processes *can* be unconscious – as shown by the Stroop effects demonstrated by Cheesman and Merikle (1984).

Bibby & Underwood argue the converse point that people can invoke volitional control when their knowledge is completely implicit. Like us, they argue that people could use "compound properties." Bibby & Underwood describe a certain higher order property that could be used to differentially apply different grammars. The knowledge cannot be completely implicit for control to occur, however; the subject must choose one of the two grammars in some way. One way of doing this is to remember a few sequences from one of the grammars and thus use the remembered sequences to activate the right knowledge (Dienes & Perner 1996). There would accordingly need to be explicit memory of specific items, even if there was implicit knowledge of the grammatical rules. Now, let us assume that the subject has, by task demands or imagination, been able to differentially activate implicit knowledge of the two grammars. We argued that measures of familiarity (e.g., RTs [reaction times] to classify) could be used to indicate whether people were using implicit knowledge or strict volitional control (based on fully explicit knowledge). Bibby & Underwood show that with a two-grammar design, RT need not predict classification performance when implicit knowledge is being used. We agree. One need not use second order effects (it is unlikely that subjects use them) to make this point; for example, subjects could think of a few g1 items to activate g1 knowledge, check the test item for g1; do the same for g2. Since the subject would do a g1 check and a g2 check each time, total RT would be the same for g1 and g2 items. On the other hand, if the subject just does a g1 check each time, RT will correlate with decisions. If the knowledge is not explicit, RT should predict classification in a one-grammar design, so this provides a way of experimentally testing the explicitness of subjects' knowledge (Buchner 1994). **Brinck** also indicates how NCC can be applied volitionally: the knowledge can be property-structure implicit, but attitude-explicit, and hence under volitional control.

Vokey & Higham suggest that the implicit/explicit distinction should be defined in terms of control rather than dissociations. We agree that control has an intimate relation to implicit/explicitness, and our article shows why there is such a relationship. We just point out that the opposition logic based on control (e.g., Jacoby 1991) is not independent of or an alternative to "dissociation logic." For Jacoby, his opposition logic can only be vindicated by dissociations; it is only by obtaining clean process dissociations that one can have confidence that the equations are the right ones for isolating different processes.

Georgieff & Rossetti describe the interesting pathologies that occur when the self is not represented as agent of the action. People suffering from schizophrenia represent themselves as observing the action (hence they are conscious of it). But they do not represent the SAS (Supervisory Attentional System) resolutions as due to the self and hence they experience the action as nonvolitional, an interesting dissociation between volition and consciousness (caused by different representations of agent of action and perceiver of action) that we had not anticipated.

R4. Visual perception

Bridgeman describes a recent experiment showing that information indexing particular objects can be effectively communicated from the ventral to the dorsal visual systems. This indicates that the sensorimotor system uses predication-explicit (factuality-implicit) representations because particular individuals were referenced (and the information was communicated between different systems). This could correspond to Bridgeman's own explanation in his closing sentence. Alternatively, the ventral system may specify a region of space (not an individual) that the dorsal system can focus on (thus illustrating how different systems can communicate when one of them deals only with predication-implicit representations, contrary to **Goldin-Meadow & Alibali**'s strict application of our claim that predication-explicitness facilitates communication beteen different systems). This is in fact the explanation given by Bridgeman and Huemer (1998).

R5. Memory

Goshen-Gottstein argues that our theory has trouble accounting for the pattern of neuropsychological dissociations, experimental dissociations, and stochastic independence observed between direct and indirect tests of memory. We would point out that one must be careful in specifying the information required and actually accessed

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

for a particular task. Goshen-Gottstein's first query concerns how we deal with a patient reported by Gabrieli et al. (1995), who had a lesion in the right occipital lobe leaving performance on direct tests intact, but performance on indirect tests impaired. Goshen-Gottstein points out that if the fully explicit representation is intact (supporting direct task performance), then all lower levels of explicitness have been represented, according to us, so there should be sufficient representations to support indirect task performance, contrary to the data. But in Gabrieli et al.'s patient, the lesion impaired only visual priming; conceptual and auditory priming, for example, were intact. This indicates that the use of representations of visual information was impaired, but not representations of the fact that certain words had been seen on a list. Thus, the direct and indirect tests relied on different facts (involving the visual makeup of a word versus the word itself, respectively), presenting no problem for our account.

Second, **Goshen-Gottstein** wonders how the independence of test performance from indirect and direct tests (e.g., Tulving et al. 1982) can fit in with our theory; he argues that according to our account p(indirect/direct) should be greater than p(indirect). However, the fact that p(indirect/direct) = p(indirect) requires just as much explanation even if one subscribes to independent memory systems to explain why explicitly stored information is not accessed by indirect procedures. Thus, we can go along and admit two physiologically separate systems (also like ventral and dorsal visual paths) or have it all in one store. The explanation must lie in independence of access for indirect and direct tests, be it to separate stores or different encodings (e.g., of different information, or the same information with or without fact markers). Anderson et al. (1998) showed how the ACT-R system, using a single interconnected set of memory chunks, can produce as much priming on indirect tests for recognised as unrecognised words. This is because the chunk storing the fact that a word was on the list can be accessed independently of the chunk relating a word and its spelling.

Finally, **Goshen-Gottstein** argues that the propositional nature of representations implies insensitivity to surface characteristics in implicit memory, yet implicit memory is highly sensitive to surface features. This is a misunderstanding of our position; there is no reason the properties represented about a stimulus and accessed by indirect tests should be restricted to the meaning of the stimulus.

Mulligan makes the related mistake of construing the difference between predication-implicitness and explicitness simply as a matter of richness of encoding. Therefore, he argues, our analysis cannot explain why elaboration during encoding affects conceptual but not perceptual priming. However, predication-implicit/explicitness is not a matter of richness of encoding. Richness of encoding is a matter of property structure (rich, articulated). Thus, one can explain the dissociation between conceptual and perceptual priming because the greater elaboration produces more conceptual primes, hence more activated material in the right part of the semantic network. Also contrary to Mulligan's claim, we do not assume that the core deficit in amnesia is a problem with conceptual-elaborative processing.

Mulligan wonders whether we will be able to experimentally establish four separate states of awareness associated with memory retrieval, given that the simple distinction between remembering and knowing may be reducible to two-criterion SDT, signal detection theory (Donaldson 1996; Hirshman & Masters 1997). However, neither of the papers cited by Mulligan as undermining the R-K distinction actually argues that the distinction should be dropped. On the contrary, both papers endorse the reality of the distinction; they just call into question *some* of the evidence for it, while approving of other evidence. We can all introspectively vindicate the difference between recollective experience and familiarity, and between volitional and nonvolitional retrieval. If you subscribe to a representational theory of mind, those different experiences must be accompanied by different representations.

R6. Development

R6.1. Infancy

Poulin-Dubois & Rakison suggest that cognitive development in infancy would be the perfect stomping ground for our theory. So let us briefly (and very speculatively) stomp that ground to show how our analysis can be applied to this field.

In the classic experiments by Baillargeon (e.g., 1987), infants of 4 months-of-age are sensitive not just to visual appearance but to deeper properties at the level of physical causality that Spelke (e.g., Spelke et al. 1995) has described as solidity, connectedness, spatio-temporal continuity, and so on. By 4½ months children also use these "Spelke properties" to individuate objects (Spelke & Kestenbaum 1986). Infants are habituated to something moving behind a left screen, and then *without* anything appearing in the spatial gap between screens, something appears from behind the right screen. Infants apparently conclude that two objects must have been involved. They dishabituate more strongly to just one object being shown at test than two. This result cannot be explained by mere feature placing (Evans 1975; target article sect. 4.1) of Spelke properties. The infants must have individuated different numbers of objects. However, they need not have explicitly predicated the Spelke properties to the identified individuals. The properties need only have been used to generate the appropriate number of individuals.

That infants may not explicitly predicate perceived properties to identified objects is suggested by a recent finding (Xu & Carey 1996) involving two clearly different types of objects: a blue rubber elephant and a red toy truck. However, the two objects move out alternately from behind a screen so that on the basis of visuo-spatial continuity there could be just one object. Not before 12 months of age do infants conclude that two objects must be involved. A possible explanation is that although the younger infants use the Spelke property of continuous spatio-temporal movement to individuate a single object, the additional properties of being red and a truck and at other times of being blue and an elephant are not predicated of that object. Hence they cannot derive a contradiction which would lead them to realise that two objects must be involved.

R6.2. Theory of mind and executive control

Sabbagh & Clegg query the interpretation of the finding by Clements and Perner (1994) in terms of implicit knowledge of false belief. In this study children listened to a story

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

about a protagonist who mistakenly thought that a desired object was at location A when in fact it had been transferred to B. At about three years most children look to A in expectation of the protagonist reappearing there, but when asked where the protagonist will reappear they point to B. This has been interpreted by Clements and Perner as showing that an implicit understanding of where the protagonist will reappear (looking in anticipation) precedes an explicit understanding (answer to question). Sabbagh & Clegg suggest – by analogy with children's difficulty with deceptive responses (Carlson et al. 1998; Russell et al. 1991) – that the young children lack the executive control to inhibit the initial predisposition to provide the usual (i.e., true) information for canonical declarative actions. It is not quite clear, once children understand that the protagonist will reappear at A, why children should have a strong "canonical" disposition to point to the wrong location, B, when asked where the protagonist will look for the object. It is also unclear why looking should be a less canonical response mode for expecting someone's reappearance (looking is not elicited as an answer to the question) than a pointing response to a question.

There are other similar possibilities, however, that pointing to B (or verbally indicating B) is not an answer to the question at all but a helpful gesture to direct the protagonist to the changed location (it is unlikely that looking would serve that purpose) and children lack the executive control to suppress these helpful pointing tendencies. This possibility is also supported by a rapidly growing literature (review by Perner & Lang, in press) showing that children's ability to respond correctly on the false belief test is linked specifically to advances in executive control. For several reasons, however, this is an unlikely explanation for the results of Clements and Perner.

In a follow up study (Clements & Perner 1997) several new conditions were used (**Ruffman** mentions further controls for this finding and **Sabbagh & Clegg**'s worry that everybody is transfixed on an implicit-explicit explanation is unwarranted). For example, children had to move a mat to where the protagonist would reappear. Children who responded spontaneously moved it correctly more often to A (as often as they looked toward A) compared to those who deliberated and moved it hesitatingly. Why should executive control fail for deliberate responses and succeed for rash ones? This result is instead compatible with the literature on dissociations between implicit and explicit knowledge.

Another reason the executive control explanation is not convincing comes from the findings of Hughes (1998) and ongoing research by Perner and Lang (1999): The strong correlation between the standard false belief task and executive control tasks is also observed for the "explanation" variant of the false belief task (Bartsch & Wellman 1989). Children observe the protagonist looking in the wrong place and have to explain why he did so. It is unlikely that children have a natural, difficult to suppress tendency to give wrong explanations or none.

An interesting question then remains: Why does understanding false belief develop in step with improvements of executive control? Perner and Lang (in press) identified several theories in the literature to explain this fact. One of them (Perner 1998) relates to our analysis of the implicit-explicit distinction. Although we did not develop this theory in the target article, **Zelazo & Frye** reconstruct it incompletely from the relevant but patchy parts in section 3.4. As they correctly point out, the theory makes use of two levels of control: contention scheduling (automatic control) and the supervisory attentional system (SAS). Norman and Shallice (1986) specify this distinction mainly in terms of a list of "SAS" tasks (novel actions, inhibition of existing habits, etc.) for which the SAS is required without specifying the particular information processing characteristics of the SAS. Perner (1998) drew on the distinction between action schemata as representational vehicles and their representational content and suggested that automatic control operates solely at the level of the vehicle, while the SAS directs control on the basis of representational content. Moreover, in order to represent these content specifications without creating confusion the SAS needs to mark them as something "desired," which requires some minimal theory of mind. As **Zelazo & Frye** correctly point out this level should be achieved at 18 months (or even earlier – a period for which there are not enough data available).

What was not mentioned in the target article (only in Perner 1998) is that for certain SAS tasks, those that require "executive inhibition" (Perner et al. 1999) a higher level of theory of mind is needed. The SAS also has to be concerned with the fact that the representational contents are carried by causally efficacious representational vehicles (i.e., the SAS needs to metarepresent the existence of action schemata as representational vehicles) in order to understand the need of inhibition: Prepotent action tendencies need to be actively inhibited because they *make* one act (causal efficacy) even though one does not want to act that way. The same understanding is required for the false belief task: a belief can *make* people look in places they really do not want to look. For that reason – according to theory – the false belief task is mastered at the same age as executive inhibition tasks such as the DCCS (dimensional change card sort) task as the data of the commentators themselves show (Frye et al. 1995).

R6.3. Gestures

Goldin-Meadow & Alibali point out that gestural expressions of reasoning processes when solving mathematical equations, for example, are not at the same level of implicitness as anticipatory eye movements in the false belief task. We agree, but perhaps for slightly different reasons. One should indeed not set visual orienting responses on a par with manual gestures, and manual gestures can indeed be put to quite different uses. In the false belief experiment the manual gesture of pointing serves as a declarative act, just like saying "there," in order to communicate the relevant information. In contrast, in solving math problems, the gestures are not intended to express or communicate anything. So we agree with the expectation for one of the proposed experiments: making speakers aware of their gestures will restrict them to expressing fact-explicit knowledge.

However, the predictions for the other experiment with the eye tracker seem less clear. The eye gaze measured in the false belief experiment is an orienting response: the child looks toward the location (A) where the protagonist mistakenly thinks the object is because the child expects the story protagonist to make an appearance there. The looking is an integral step in how the child interacts with the story events. Without looking toward A the child would not see

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

the protagonist. In contrast, gestures accompanying math problems are not integral in the same sense. As **Goldin-Meadow & Alibali** point out, they are "symbolic," certainly in the sense that they map a thought process without being part of that process. The problem with these commentators' prediction about eye gaze patterns is that in the context of the math problems eye gaze might serve the same purpose as manual gestures of mapping thought processes without being an integral part of them. Hence, eye gaze patterns in this task may be as predication-explicit as manual gestures according to these commentators' argument.

Part of Goldin-Meadow & Alibali's argument as to why the knowledge underlying gestures is predication-explicit rests on the observation that knowledge expressed in gestures is often also used for generated solutions later. This can be explained as a piece of knowledge getting hold of different response modalities. This kind of "accessibility," however, does not require predication-explicitness. Explicit predication is required when one part of the system is *looking* for a particular kind of information that exists in another functionally unrelated part. Hence, the more relevant evidence for predication-explicitness is that children rate solutions that conform to their gesturally expressed knowledge as more reliable than their solutions that conform to unexpressed procedures. However, these confidence ratings suggest that gesturally expressed knowledge is not only predication-explicit but also factuality-explicit, that is, there is some awareness of the gesturally-expressed knowledge being reliable to some degree. This underlines the contrast with anticipatory looking in the false belief task according to the data of **Ruffman:** Children rate the solution expressed by their anticipatory looking as having zero probability (Ruffman et al. 1998).

Alibali & Koedinger wonder in this context, what advantage accrues from thinking about procedural knowledge as a "fact." This question smacks of a misreading of what we mean by these terms. We are in no way concerned about whether or not the existence of some procedural knowledge is a fact. Rather, we are concerned about whether knowledge (embedded in procedures or otherwise) represents (see sect. R1.5 of this response for refinement) a fact *as* a fact or not. For example, there is the fact that for $y = 2x + 1$ and $x = 5$, y equals 11. A calculator knows this purely procedurally. Given the equation and a value for x, it will spit out "11." It does not know that this underlying regularity is a fact, that is, it could not pretend that the answer is 12, or provide a confidence rating for the answer "11." That is one reason why the ability of adults to provide higher confidence ratings for solutions conforming to gestured procedures than for solutions conforming to ungestured procedures indicates factuality-explicit knowledge.

Moreover, **Alibali & Koedinger** suggest that the findings on gestured versus verbalised knowledge of procedures can be satisfactorily modelled within ACT-R by the differential activation of declarative chunks. Weakly activated chunks may fail to fire complex language productions but may fire simpler, more well-practised productions for gestures. This suggestion strikes us as odd for the following reasons:

1. Does the fact that new and better knowledge is often only expressed in gestures mean that children get a lot more practice in gesturing algebraic procedures than in talking about them?

2. How does this modelling square with **Goldin-Meadow & Alibali**'s suggestion that knowledge expressed in gestures is factuality-implicit? Is the suggestion that strength of activation represents factuality: Does being above a certain level of activation represent that it is a fact? If that is so, then how does ACT-R implement a well practised procedure that is deemed unreliable, or an overlearned procedure that was once explicit and then, through automatisation, has become implicit and unverbalisable again? The latter is a well known problem for threshold theories of consciousness (Baars & McGovern 1996, p. 76).

R7. Models of learning

Jiménez & Cleeremans imply that knowledge need not be representational: Tuning relevant neural pathways does not form a representation, it is just a process; the corresponding claim in a connectionist network is that weights do not represent, only activation patterns do. But in our functional definition of representation, weights are representational, because they have the function of covarying with various structures in the world (remember that RTK does not imply a language of thought). In a Hebbian network, a weight linking two nodes (representing, say, the presence of A and B, respectively) has the function of indicating the covariation between A and B. That is, the weight represents that covariation. Correspondingly, the weight has all the features of a representation (Perner 1991): It is singular (it is about A and B and not C or D); it can misrepresent (for example, if the nodes themselves misrepresented A or B by being triggered by a C or a D on a dark night, then the Hebbian rule would lead the weights to likewise misrepresent), and so on. The weight has this content, but it is nonconceptual content: It is not composed of constituents that satisfy the generality constraint; nor does it satisfy the generality constraint itself (and *typically* activation patterns carry NCC as well). Nonetheless, weights and neural pathways are representational (on a teleological account, and teleological views of representation are perhaps the philosophically dominant ones). As long as Jiménez & Cleeremans accept that neural pathways have certain functions (of indicating certain contents), our framework remains applicable to the priming cases they mention.

Vokey & Higham consider other learning mechanisms – for example, the storage of instances – for which they question our use of the term implicit. Instead, such mechanisms produce what might be better described as latent knowledge, distributed over the data base. We agree that the knowledge latent in exemplars is not ipso facto implicit in our sense. But we think it is the way in which the knowledge is implicit, rather than simply latent, that captures an important part of the attraction of paradigms like artificial grammar learning.[2] The exemplars or episodes may be implicit (not predicated of a particular spatiotemporal learning context) or explicit (capable of providing recollective experience). Inferences based on the exemplars (in producing classification decisions, for example) may likewise be explicit or implicit; that is, represented as knowledge because their appropriate causal origin is represented, or considered as mere guesses. Implicit learning, according to most people's intuitions, would be said to occur when either the exemplars themselves or the inferences based on them are implicit (in our sense), not simply latent (in Vokey & Higham's sense). If the knowledge were simply latent, it would leave open the possibility that people could describe

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press

which training items they brought to mind (recollective experience) and how they assessed their similarity with the test item (justified knowledge of their grammaticality judgements).

Marescaux & Chambres indicate the complexity of the artificial grammar learning task and the range of learning mechanisms (connectionist, instance storage, etc.) that may be responsible for performance. They correctly point out (as did we in the target article) that confidence judgements about grammaticality judgements do not provide direct evidence about the implicit/explicit status of knowledge of the grammar per se. For the latter, we need to infer how the knowledge is represented. Agreed, this makes life difficult, but exactly the same problem exists in inferring the implicit/explicit nature of the knowledge whether one subscribes to our framework or not. If one can plausibly infer what the "rules" are (instances, n-grams, etc.), then our framework enables one to test in what ways the knowledge is implicit or explicit.

Various commentators recommend the use of the ACT-R (Anderson & Lebiere 1998) framework for understanding implicit knowledge. The accomplishments of ACT-R are indeed impressive. There is no apparent inconsistency between the ACT-R model and our framework, and points of concordance are noted by **Lebiere & Wallach.** However, whether ACT-R is able to incorporate fully those distinctions made by our framework that are necessary for understanding human cognition remains to be seen. For example, the account of explicit recognition memory (Anderson et al. 1998; see sect. R5, Memory) makes use of a context chunk that represents particular words (e.g., "hare") as members of the learned word list. Lebiere & Wallach consider this an instance of explicit predication but the model leaves open whether the context chunk represents the complex property, "list with 'hare' in it," or the predicating proposition, "The list has 'hare' in it." Apart from the model users' intentions, what makes the context chunk a representation of the latter rather than the former? How would the model distinguish these two psychologically different cases?

Alibali & Koedinger even suggest that ACT-R leads to predictions at odds with our theory: Our theory cannot easily interpret a person being able to state a theorem but being unable to apply the rule in context. This is not difficult for us – according to any account, to apply the theorem the person must (a) realise its relevance; and (b) have other supporting knowledge relevant to the problem set. The person may be lacking (a) or (b), even if knowledge of the theorem is quite explicit. Alibali & Koedinger further wonder how our theory accounts for different types of implicitness observed in people. In their examples of different degrees of implicitness, there is a need to distinguish the generality of the rule induced from its explicitness (i.e., a more general rule does not eo ipso mean a more explicit one), a distinction often missed in the literature. For example, in their second paragraph, the number you add to the first number to get the second increases by one in each successive number pair. This rule would be difficult to apply to pairs much smaller or larger than the pairs trained on, but there need not be anything more implicit about the rule (in our sense) than the rule $y = 2x + 1$.

Noelle argues that the distinction between implicit and explicit learning may be best understood in terms of different brain systems rather than different propositional attitudes. We agree that the sources of dissociations within a knowledge domain are unlikely to be due exclusively to content differences. In many cases, different brain regions are likely involved. However, the different brain regions can compute different contents, which gives them their implicit or explicit function. Noelle does not confront the question of why we call the knowledge in the different systems implicit or explicit; this is where our framework clarifies. Thus, we can explain why completely different brain regions show similar dissociations, for example, vision (parietal and temporal cortex) versus theory of mind (prefrontal cortex).

Finally, **Gorman** considers the special case of learning involved in scientific discovery. He says he has argued that Bell followed an "implicit confirmation heuristic" (Gorman 1995). We are not entirely sure what was meant to be implicit. Bell of course does not say that he is following a confirmation heuristic in his notebooks; but he may just have regarded this as not something useful to put in his notebook. Similarly, for protocol analyses; they provide suggestive but not definitive evidence about which heuristics may be implicit, because subjects will only say in their protocol what they think the experimenter is interested in hearing. What was left out could perhaps be confidently reported if the subject were asked directly about it (the normal problems with relying on free report as an exhaustive measure of explicit knowledge). Nonetheless, there is plenty of scope for interesting further work on the role of implicit knowledge in scientific discovery.

R8. Conclusion

The different views with which the commentators confronted us have greatly expanded and clarified our own explicit understanding of the implications implicit in our ideas. It is reassuring that our ideas stand up to such insightful scrutiny, and we look forward to their further development. Our final comment is for the Zen-like commentary of M. J. **O'Brien.** In response, we merely raise a finger. If O'Brien raises a finger back, we will chop it off. And in that moment he will attain the attitude of enlightenment.

ACKNOWLEDGMENTS

We thank Ingar Brinck and Ron Chrisley for guidance through the mysteries of nonconceptual content, Ron Chrisley for other useful discussions, and Bruce Bridgeman for making his data available to us.

NOTES

1. A fuller exposition would make clear that the structure-implicit facts are those that define the conditions that must hold only in *nearby* possible worlds for the representation to have the meaning it has. Facts like laws of physics, chemistry, or biology must hold even in relatively distant possible worlds; if they did not hold, the world, and not just the meaning of a few representations, would be completely different. (Thanks to Ron Chrisley for making this suggestion.)

2. Being latent captures another, separate part of the attraction!

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively

Aglioti, S., DeSouza, J. F. X. & Goodale, M. A. (1995) Size-contrast illusions deceive the eye but not the hand. *Current Biology* 5(6):679–85. [aZD]

Ahmed, A. & Ruffman, T. (1998) Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task? *Developmental Psychology* 34:441–53. [TR]

Alibali, M. W (1999) How do children change their minds? Strategy change can be gradual or abrupt. *Developmental Psychology* 35:127–45. [MWA]

Alibali, M. W., Bassok, M., Olseth, K. L., Syc, S. E. & Goldin-Meadow, S. (1999) Illuminating mental representations through speech and gesture. *Psychological Science* 10:327–33. [SG-M]

Alibali, M. W. & Goldin-Meadow, S. (1993) Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology* 25:468–523. [SG-M]

Alibali, M. W., McNeil, N. M. & Perrott, M. A. (1998) What makes children change their minds? Changes in encoding lead to changes in strategy selection. In: *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society,* ed. M. A. Gernsbacher & S. Derry. Erlbaum. [SG-M]

Allen, S. W. & Brooks, L. R. (1991) Specializing the operations of an explicit rule. *Journal of Experimental Psychology: General* 120:1–19. [JRV]

Anderson, J. R. (1976) *Language, memory and thought.* Erlbaum. [aZD]

Anderson, J. R., Bothell, D., Lebiere, C. & Matessa, M. (1998) An integrated theory of list memory. *Journal of Memory and Language* 38(4):341–80. [CL, rJP]

Anderson, J. R. & Lebiere, C. (1998) *The atomic components of thought.* Erlbaum. [MWA, RAC, CL, rJP, NAT]

Armstrong, D. (1980) *The nature of mind and other essays.* Cornell University Press. [aZD]

Baars, B. J. & McGovern, K. (1996) Cognitive views of consciousness: What are the facts? How can we explain them? In: *The science of consciousness: Psychological, neuropsychological and clinical reviews,* ed. M. Velmans. Routledge. [rJP]

Baddeley, A. (1986) Modularity, mass-action and memory. Special Issue: Human memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 38:527–33. [aZD]

Badgaiyan, R. D. & Posner, M. I. (1997) Time course of cortical activations in implicit and explicit recall. *Journal of Neuroscience* 17:4904–13. [EM]

Baillargeon, R. (1987) Object permanence in 3½ and 4½-month-old infants. *Developmental Psychology* 23:655–64. [rJP, SG-M]

Bargh, J. (1992) The ecology of automaticity: Towards establishing the conditions needed to produce automatic processing effect. *American Journal of Psychology* 105:181–99. [JT]

Bartsch, K. & Wellman, H. M. (1989) Young children's attribution of action to beliefs and desires. *Child Development* 60:946–64. [rJP]

Barwise, J. (1987) Unburdening the language of thought. *Mind and Language* 2:82–96. [aZD]

Barwise, J. & Perry, J. (1983) *Situations and attitudes.* MIT Press. [aZD]

Bauer, G. H. & Johnson, C. M. (1994) Trained motor imitation by bottlenose dolphins (*Tursiops truncatus*). *Perceptual and Motor Skills* 79:1307–15. [EM]

Bechtel, W. & Abrahamsen, A. (1991) *Connectionism and the mind: An introduction to parallel processing in networks.* Blackwell. [aZD, MEG]

Bermudez, J. L. (1995) Nonconceptual content: From perceptual experience to subpersonal computational states. *Mind and Language* 10:333–69. [rJP]

Berry, D. C., ed. (1997) *How implicit is implicit learning?* Oxford University Press. [aZD]

Berry, D. C. & Dienes, Z. (1993) *Implicit learning: Theoretical and empirical issues.* Erlbaum. [P-JM]

Berry, D. C. & Broadbent, D. E. (1988) Interaction tasks and the implicit-explicit distinction. *British Journal of Psychology* 79:251–72. [DCN]

Bertenthal, B. I. (1993) Infants' perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In: *Visual perception and cognition in infancy: Carnegie Mellon Symposium on Cognition,* ed. G. Granrud. Erlbaum. [DP-D]

(1996) Origins and early development of perception, action, and representation. *Annual Review of Psychology* 47:431–59. [SG-M]

Block, N. (1994) Consciousness. In: *A companion to the philosophy of mind,* ed. S. Guttenplan. Blackwell. [aZD]

(1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18(2):227–87. [aZD, rJP]

Bornstein, R. F. (1989) Exposure and affect: Overview and meta-analysis of research 1968–1987. *Psychological Bulletin* 106:265–89. [aZD]

(1992) Subliminal mere exposure effects. In: *Perception without awareness: Cognitive, clinical, and social perspectives,* ed. R. F. Bornstein & T. S. Pittman. Guilford Press. [RFB]

(1998) Implicit and self-attributed dependancy strivings: Differential relationships to laboratory and field measures of help-seeking. *Journal of Personality and Social Psychology* 75:778–87. [RFB]

Bornstein, R. F., Rossner, S. C., Hill, E. L. & Stepanian, M. L. (1994) Face validity and fakeability of objective and projective measures of dependency. *Journal of Personality Assessment* 63:363–86. [RFB]

Boucher, J. (1998a) Time processing in human behaviour and evolution. Poster presented at the Hang Seng International Conference on the Evolution of Mind, Sheffield, United Kingdom, June 1998. [JB]

(1998b) Time parsing, normal language acquisition, and specific language impairments. Paper presented at the Child Language Seminar, Sheffield, UK, September 1998. [JB]

(in press a) Lost in a sea of time: Time parsing and autism. In: *Time and memory,* ed. T. McCormack & C. Hoerl. Oxford University Press. [JB]

(in press b) Time parsing, normal language acquisition, and language-related developmental disorders. In: *New directions in research into language development and disorders,* ed. M. Perkins & S. Howard. Plenum Press. [JB]

Bridgeman, B. (1991) Complementary cognitive and motor image processing. In: *Presbyopia research: From molecular biology to visual adaptation,* ed. G. Obrecht & L. W. Stark. Plenum Press. [BB, aZD, SG-M]

(1992) Conscious vs. unconscious processes. The case of vision. *Theory and Psychology* 2:73–88. [NG]

Bridgeman, B. & Huemer, V. (1998) A spatially oriented decision does not induce consciousness in a motor task. *Consciousness and Cognition* 7:454–64. [BB, rJP]

Bridgeman, B., Kirch, M. & Sperling, A. (1981) Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics* 29:336–42. [BB]

Bridgeman, B., Peery, S. & Anand, S. (1997) Interaction of cognitive and sensorimotor maps of visual space. *Perception and Psychophysics* 59:456–69. [BB, aZD, SG-M]

Brinck, I. (1997) The indexical "I." Kluwer. [rJP]

Brooks, L. R. (1978) Non-analytic concept formation and memory for instances. In: *Cognition and Concepts,* ed. E. Rosch & B. Lloyd. Erlbaum. [JRV]

Brooks, L. R., Vokey, J. R. & Higham, P. A. (1997) Two bases for similarity judgments with a category. In: *Proceedings of Simacat97: An interdisciplinary workshop on similarity and categorization.* Edinburgh, Scotland. [JRV]

Brooks, L. R. & Wood, T. (1997) Identification in service of use: Characteristic of every- day concept learning. In: *Abstracts of the Psychonomic Society: 38th Annual Meeting.* The Psychonomic Society. [DCN]

Brown, G. D. A. & Vousden, J. (1998) Adaptive analysis of sequential behaviour: Oscillators as rational mechanisms. In: *Rational models of cognition,* ed. M. Oaksford & N. Chater. Oxford University Press. [JB]

Buchner, A. (1994) Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:550–66. [aZD, rJP]

Buckner, R. L. & Koutstaal, W. (1998) Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proceedings of the National Academy of Sciences, USA* 95:891–98. [EM]

Byrne, R. (1995) *The thinking ape: Evolutionary origins of intelligence.* Oxford University Press. [rJP]

Byrne, R. W. & Whiten, A., eds. (1988) *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans.* Clarendon Press. [AC-M]

Campbell, J. (1993) The role of physical objects in spatial thinking. In: *Spatial representation,* ed. N. Eilan, R. McCarthy & B. Brewer. Blackwell. [aZD]

Carlson, R. A. (1997) *Experienced cognition.* Erlbaum. [RAC]

Carlson, S. M., Moses, L. J. & Hix, H. R. (1998) The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development* 69(3):672–91. [rJP, MAS]

Carruthers, P. (1992) Consciousness and concepts. *Proceedings of the Aristotelian Society, Supplementary Vol. LXVI:*42–59. [aZD]

(1996) *Language thought and consciousness: An essay in philosophical psychology.* Cambridge University Press. [aZD, GO, PDZ]

Carstairs-McCarthy, A. (1998) Synonymy avoidance, phonology and the origin of syntax. In: *Approaches to the evolution of language: Social and cognitive bases,* ed. J. R. Hurford, M. Studdert-Kennedy & C. Knight. Cambridge University Press. [AC-M]

(1999) *The origins of complex language: An inquiry into the evolutionary beginnings of sentences, syllables and truth.* Oxford University Press. [AC-M]

Chadwick, P. & Birchwood, M. (1994) The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry* 164:190–201. [NG]

Chan, C. (1992) Implicit cognitive processes: Theoretical issues and applications in

computer systems design. Unpublished D. Phil. thesis, University of Oxford. [aZD]
Chater, N. (1997) Simplicity and the mind. *The Psychologist* 10:495–98. [GOv]
Cheesman, J. & Merikle, P. M. (1984) Priming with and without awareness. *Perception and Psychophysics* 36(4):387–95. [aZD, rJP, CEW]
(1986) Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology* 40(4):343–67. [aZD]
Cheney, D. L. & Seyfarth, R. M. (1990) *How monkeys see the world: Inside the mind of another species.* University of Chicago Press. [AC-M]
Chomsky, N. (1981) *Lectures on government and binding.* Foris. [PMP]
(1986a) *Knowledge of language.* Praeger. [PMP]
(1986b) Changing perspectives on knowledge and the use of knowledge. In: *The representation of knowledge and belief: Arizona Colloquium in Cognition.* University of Arizona Press. [MAS]
Chrisley, R. L. (1996) Non-conceptual psychological explanation: Content and computation. D. Phil. thesis, University of Oxford. [rJP]
Church, R. B. & Goldin-Meadow, S. (1986) The mismatch between gesture and speech as an idex of transitional knowledge. *Cognition* 23:43–71. [aZD, SG-M]
Clark, A. & Karmiloff-Smith, A. (1993) The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language* 8:487–519. [aZD]
Cleeremans, A. (1997) Principles for implicit learning. In: *How implicit is implicit learning?,* ed. D. C. Berry. Oxford University Press. [aZD, LJ]
Cleeremans, A., Destrebecqz, A. & Boyer, M. (1998) Implicit learning: News from the front. *Trends in Cognitive Sciences* 2:406–16. [GOv]
Cleeremans, A. & Jiménez, L. (submitted) Implicit cognition with the symbolic metaphor of mind: Theoretical and methodological issues. [LJ]
Clements, W. A. (1995) Implicit theories of mind. Unpublished doctoral dissertation, University of Sussex. [aZD]
Clements, W. A. & Perner, J. (1994) Implicit understanding of belief. *Cognitive Development* 9:377–97. [MWA, aZD, rJP, SG-M, DP-D, TR, MAS]
(1996) Implicit understanding of belief at three in action. Unpublished manuscript, University of Sussex. [SG-M, MAS]
(1997) When actions really do speak louder than words but only implicitly: Young children's understanding of false belief in action. Unpublished manuscript, University of Sussex. [aZD, rJP, NG]
Clements, W. A., Rustin, C. & McCallum, S. (1997) Promoting the transition from implicit to explicit understanding: A training study of false belief. Unpublished manuscript, University of Sussex. [aZD]
Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J. & Cohen, G. M. (1997) Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General* 126:393–413. [aZD, NWM]
Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276. [aZD]
Cowey, A. & Stoerig, P. (1995) Blindsight in monkeys. *Nature* 373:247–49. [SN, rJP]
Cummins, R. (1986) Inexplicit representation. In: *The representation of knowledge and belief,* ed. M. Brand & R. Harnish. University of Arizona Press. [GO]
Curran, T. & Keele, S. W. (1993) Attentional and non-attentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:189–202. [CL]
Currie, G. (1982) *Frege, an introduction to his philosophy.* Harvester Press. [aZD]
Currie, G. & Ravenscroft, I. (in press a) *Meeting of minds: Thought, perception and imagination.* Oxford University Press. [aZD]
(in press b) The development of pretense. In: *Meeting of minds: Thought, perception and imagination,* ed. G. Currie & I. Ravenscroft. Oxford University Press. [rJP]
Cussins, A. (1992) Content, embodiment and objectivity: The theory of cognitive trails. *Mind* 101:651–88. [rJP]
Dagenbach, D., Carr, T. H. & Wilhelmsen, A. (1989) Talk-induced strategies and near- threshold priming: Conscious influences on unconscious perception. *Journal of Memory and Language* 28:412–43. [aZD]
Davidson, D. (1963) Actions, reasons, and causes. *Journal of Philosophy* 60:685–700. [aZD]
Debner, J. A. & Jacoby, L. L. (1994) Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:304–17. [aZD]
Dennett, D. C. (1978) *Brainstorms.* Bradford. [aZD]
(1982) Styles of mental representation. *Proceedings of the Aristotelian Society New Series* 83:213–26. [GO]
Denny, E. B. & Hunt, R. R. (1992) Affective valence and memory in depression: Dissociation of recall and fragment completion. *Journal of Abnormal Psychology* 101:575–80. [NWM]
DeYoe, E. A. & Van Essen, D. C. (1988) Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences* 11:219–26. [CEW]
Diamond, A. (1985) Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development* 56:868–83. [rJP]
Diamond, A. & Goldman-Rakic, P. S. (1989) Comparison of human infants and infant rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research* 74:24–40. [aZD]
Dienes, Z. (1992) Connectionist and memory array models of artificial grammar learning. *Cognitive Science* 16:41–79. [aZD]
Dienes, Z. & Altmann, G. (1997) Transfer of implicit knowledge across domains? How implicit and how abstract? In: *How implicit is implicit learning?,* ed. D. Berry. Oxford University Press. [aZD]
Dienes, Z., Altmann, G. T. M., Kwan, L. & Goode, A. (1995) Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:1322–38. [PAB, aZD]
Dienes, Z. & Berry, D. (1997) Implicit learning: Below the subjective threshold. *Psychonomic Bulletin and Review* 4:3–23. [aZD]
Dienes, Z., Broadbent, D. E. & Berry, D. (1991) Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:875–78. [GO]
Dienes, Z., Kurz, A., Bernhaupt, R. & Perner, J. (1997) Application of implicit knowledge: Deterministic or probabilistic? *Psychologica Belgica* 37:89–112. [aZD]
Dienes, Z. & Perner, J. (1996) Implicit knowledge in people and connectionist networks. In: *Implicit cognition,* ed. G. Underwood. Oxford University Press. [aZD, rJP]
Dokic, J. (1997) Two meta-representational theories of episodic memory. Paper presented at the Annual Meeting of the European Society for Philosophy in Psychology in Padua, Italy, August 1997. [aZD]
Donaldson, W. (1996) The role of decision processes in remembering and knowing. *Memory and Cognition* 24:523–33. [NWM, rJP]
Dretske, F. (1988) *Explaining behavior: Reasons in a world of causes.* MIT Press. [aZD]
(1995) *Naturalizing the mind.* MIT Press. [aZD]
Dulany, D. E. (1991) Conscious representation and thought systems. In: *Advances in social cognition, vol. 4,* ed. R. S. Wyer & T. K. Srull. Erlbaum. [aZD, rJP, JT]
(1997) Consciousness in the explicit (deliberative) and implicit (evocative). In: *Scientific approaches to the study of consciousness,* ed. J. D. Cohen & J. W. Schooler. Erlbaum. [aZD, rJP]
Dulany, D. E., Carlson, R. C. & Dewey, G. I. (1984) A case of syntactical learning and judgement: How conscious and how abstract? *Journal of Experimental Psychology: General* 113:541–55. [P-JM, GO]
Dunbar, K. (1995) How scientists really reason: Scientific reasoning in real-world laboratories. In: *The nature of insight,* ed. R. J. Sternberg & J. Davidson. MIT Press. [MEG]
(1997) How scientists think. In: *Creative thought,* ed. T. B. Ward, S. M. Smith & J. Vaid. American Psychological Association. [MEG]
Dunbar, R. (1996) *Grooming, gossip and the evolution of language.* Faber and Faber. [AC-M]
Dwyer, S. & Pietroski, P. (1996) Believing in language. *Philosophy of Science* 63:38–73. [PMP]
Ebbinghaus, H. (1885) *Über das Gedächtnis.* Duncker und Humblot. [rJP]
Eichenbaum, H. (1997) Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology* 48:547–72. [EM]
Elman, J. (1990) Finding structure in time. *Cognitive Science* 14:178–211. [JB]
Engelkamp, J. & Wippich, W. (1995) Current issues in implicit and explicit memory. *Psychological Research* 57:143–55. [EM]
Ericsson, K. A. & Simon, H. A. (1984) *Protocol analysis: Verbal reports as data.* MIT Press. [MEG]
Eriksen, C. W. (1960) Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review* 67:279–300. [aZD]
Evans, G. (1975) Identity and predication. *The Journal of Philosophy* 72(13):343–63. [aZD, rJP]
Evans, J. St. B. T. (in press) What could and could not be a strategy in reasoning. In: *Deductive reasoning and strategies,* ed. W. Schaeken, A. Vandierendonck & G. De Vooght. Erlbaum. [JSBTE]
Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning.* Psychology Press. [JSBTE]
(1997) Rationality in reasoning: The problem of deductive competence. *Current Psychology of Cognition* 16:3–38. [JSBTE]
Feist, G. & Gorman, M. E. (1998) The psychology of science: Review and integration of a nascent discipline. *Review of General Psychology* 2(1):3–47. [MEG]

Field, H. (1978) Mental representation. *Erkenntenis* 13:9–61. [aZD]
Fodor, J. A. (1975) *The language of thought.* Harvard University Press. [rJP]
(1978) Propositional attitudes. *The Monist* 61:501–23. [IB, aZD]
(1983) *The modularity of mind.* MIT Press. [aZD]
(1987a) A situated grandmother? Some remarks on proposals by Barwise and Perry. *Mind and Language* 2:64–81. [aZD]
(1987b) Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In: *Modularity in knowledge representation and natural-language understanding,* ed. J. L. Garfield. MIT Press. [aZD]
(1987c) *Psychosemantics.* MIT Press. [SN]
(1998) *Concepts: Where cognitive science went wrong.* Clarendon Press. [JAH]
Fourneret, P. Jeannerod, M. (1998) Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 36(11):1133–40. [NG]
Fowler, C. A., Wolford, G., Slade, R. & Tassinary, L. (1981) Lexical access with and without awareness. *Journal of Experimental Psychology: General* 110:341–62. [aZD]
Fraisse, P. (1963) *The psychology of time,* trans. J. Leith. Harper & Row. [BB]
Frege, G. (1977) *Logical investigations.* Blackwell. [LS]
(1979) *Posthumous papers.* Blackwell. [LS]
(1980) Translations from *The philosophical writings of Gottlieb Frege, 3rd edition,* ed. P. Geach & M. Black. Cornell University Press. [AC-M]
Frensch, P. A. (1998) One concept, multiple meanings: On how to define the concept of implicit learning. In: *Handbook of implicit learning,* ed. M. A. Stadler & P. A. Frensch. Sage. [P-JM]
Frith, C. D. (1992) *The neuropsychology of schizophrenia.* Erlbaum. [NG]
(1995) Consciousness is for other people. *Behavioral and Brain Sciences* 18:682–83. [NG]
Frye, D., Zelazo, P. D. & Palfai, T. (1995) Theory of mind and rule-based reasoning. *Cognitive Development* 10:483–527. [rJP, PDZ]
Gabrieli, J. D. E., Fleishman, D. A., Keane, M. M., Reminger, S. L. & Morrell, F. (1995) Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science* 6:76–82. [YG-G, rJP]
Gallistel, C. R. (1990) *The organisation of learning.* MIT Press. [JB]
Garber, P., Alibali, M. W. & Goldin-Meadow, S. (1998) Knowledge conveyed in gesture is not tied to the hands. *Child Development* 69:75–84. [MWA, SG-M]
Gardiner, J. (1988) Functional aspects of recollective experience. *Memory and Cognition* 16:309–13. [aZD]
Gardiner, J. M., Ramponi, C. & Richardson-Klavehn, A. (1998) Experiences of remembering, knowing and guessing. *Consciousness and Cognition* 7:1–26. [NWM]
Gelman, R., Durgin, F. & Kaufman, L. (1995) Distinguishing between animates and inanimates: Not by motion alone. In: *Causal cognition: A multidisciplinary debate,* ed. D. Sperber, D. Premack & A. J. Premack. Clarendon Press. [DP-D]
Gentilucci, M., Chieffi, S. & Daprati, E. (in press) Visual illusion and action. *Neuropsychologia.* [aZD]
Georgieff, N. & Jeannerod, M. (1998) Beyond consciousness of external reality: A "who" system for consciousness of action and self-consciousness. *Consciousness and Cognition* 7:465–78. [NG]
Gewei, Y. & van-Raaij, F. W. (1997) What inhibits the mere-exposure effect: Recollection or familiarity? *Journal of Economic Psychology* 18:629–48. [aZD]
Gibson, J. J. (1950) *The perception of the visual world.* Houghton-Mifflin. [aZD]
Goldin-Meadow, S. (1997) When gesture and words speak differently. *Current Directions in Psychological Science* 6:138–43. [SG-M]
Goldin-Meadow, S., Alibali, M. W. & Church, R. B. (1993) Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review* 100:279–97. [MWA, aZD, SG-M]
Goodale, M. A. & Milner, A. D. (1992) Separate visual pathways for perception and action. *Trends in the Neurosciences* 15:20–25. [NG]
Goodall, J. (1986) *The chimpanzees of Gombe: Patterns of behavior.* Belknap Press. [AC-M]
Gopnik, A. (1993) How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16:1–113. [aZD]
Gordon, R. M. (1995) Simulation without introspection or inference from me to you. In: *Mental simulation: Evaluations and applications,* ed. M. Davies & T. Stone. Blackwell. [aZD, rJP]
Gorman, M. E. (1995) Confirmation, disconfirmation, and invention: The case of Alexander Graham Bell and the telephone. *Thinking and Reasoning* 1(1):31–53. [rJP, MEG]
(1998) *Transforming nature: Ethics, invention and design.* Kluwer Academic. [MEG]
Greenwald, A. G. (1992) New Look 3: Unconscious cognition reclaimed. *American Psychologist* 47(6):766–79. [aZD]
Gusterson, H. (1996) *Nuclear rites: A weapons laboratory at the end of the cold war.* University of California Press. [MEG]
Güzeldere, G. (1995) Is consciousness the perception of what passes in one's own mind? In: *Conscious experience,* ed. T. Metzinger. Schöningh. [aZD]
Haverty, L., Koedinger, K. R., Klahr, D. & Alibali, M. W. (1999) Solving inductive reasoning problems in mathematics: Not-so-trivial PURSUIT. *Cognitive Science.* (in press). [MWA]
Hayes, S. C. (1992) Verbal relations, time, and suicide. In: *Understanding verbal relations,* ed. S. C. Hayes & L. J. Hayes. Context Press. [GOv]
Hayman, C. A. G. & Tulving, E. (1989) Contingent dissociations between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:228–40. [YG-G]
Heyes, C. & Dickinson, A. (1993) The intentionality of animal action. In: *Consciousness: Psychological and philosophical essays,* ed. M. Davies & G. W. Humphreys. Blackwell. [aZD, rJP]
Higham, P. A. & Brooks, L. R. (1997) Learning the experimenter's design: Tacit sensitivity to the structure of memory lists. *The Quarterly Journal of Experimental Psychology* 50A:199–215. [JRV]
Higham, P. A. & Vokey, J. R. (1999) The controlled application of a strategy can still produce automatic effects. (in preparation). [JRV]
Higham, P. A., Vokey, J. R. & Pritchard, J. L. (in press) Beyond task dissociations: Evidence for controlled and automatic decisions in artificial grammar learning. *Journal of Experimental Psychology: General.* [JRV]
Hirschman, E. & Henzler, A. (1998) The role of decision processes in conscious recollection. *Psychological Science* 9:61–65. [NWM]
Hirshman, E. & Master, S. (1997) Modelling the conscious correlates of recognition memory: Reflections on the Remember-Know paradigm. *Memory and Cognition* 25(3):345–51. [NWM, rJP]
Holender, D. (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences* 9:1–66. [aZD, JT]
Homer, B. D. & Olson, D. R. (1999) Literacy and children's conception of words. *Written Language and Literacy* 2(1):113–37. [BDH]
Hughes, C. (1998) Executive functions in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology* 16(2):233–53. [rJP]
Hume, D. (1969) *A treatise of human nature.* Penguin. (Original work published 1739–1740). [GOv]
Jacoby, L. L. (1991) A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language* 30:513–41. [aZD, SK, rJP, JRV]
Jacoby, L. L. & Dallas, M. (1981) On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General* 110:306–40. [aZD, YG-G]
Jacoby, L. L., Levy, B. & Steibach, K. (1992) Episodic transfer and automaticity: Integration of data-driven and conceptually driven processing in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:15–24. [JT]
Jacoby, L. L., Lindsay, D. S. & Toth, J. P. (1992) Unconscious influences revealed: Attention, awareness, and control. *American Psychologist* 47:802–809. [aZD]
Jacoby, L. L., Toth, J. P. & Yonelinas, A. P. (1993) Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General* 122:139–54. [SK]
James, W. (1890) *The principles of psychology.* Macmillan. [rJP]
Jeannerod, M. (1994) The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences* 17:187–245. [NG]
(1997) *The cognitive neuroscience of action.* Blackwell. [NG]
Jeannerod, M. & Rossetti, Y. (1993) Visuomotor coordination as a dissociable function: Experimental and clinical evidence. In: *Visual perceptual defects: Baillière's clinical neurology, international practise and research,* ed. C. Kennard. Baillière Tindall. [NG]
Jiménez, L. (1997) Implicit learning: Conceptual and methodological issues. *Psychologica Belgica* 37:9–28. [P-JM]
Johnson-Laird, P. N. & Byrne, R. (1991) *Deduction.* Erlbaum. [JSBTE]
Kant, I. (1933) *Critique of pure reason.* Macmillan. [LS]
Karmiloff-Smith, A. (1986) From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition* 23:95–147. [MWA, aZD]
(1992) *Beyond modularity: A developmental perspective on cognitive science.* MIT Press. [MWA, aZD, BDH, rJP, DP-D]
Keil, F. C. (1989) *Concepts, kinds, and cognitive development.* MIT Press/A Bradford Book. [rJP]
(1998) The most basic units of thought do more, and less, than point. *Behavioral and Brain Sciences* 21:75–76. [rJP]
Kihlstrom, J. F. (1996) Perception without awareness of what is perceived, learning without awareness of what is learned. In: *The science of consciousness:*

Psychological, neuropsychological and clinical reviews, ed. M. Velmans. Routledge. [aZD]

Kihlstrom, J. F., Barnhardt, T. & Tataryn, D. (1992) Implicit perception. In: *Perception without awareness: Cognitive, clinical, and social perspectives,* ed. R. Bornstein & T. Pittman. Guilford Press. [aZD]

Kirsh, D. (1991) When is information explicitly represented? In: *Information, thought, and content,* ed. P. Hanson. UBC Press. [aZD]

Kirsner, K., Dunn, J. C. & Standen, P. (1989) Domain-specific resources in word recognition. In: *Implicit memory: Theoretical issues,* ed. S. Lewandowsky, J. C. Dunn & K. Kirsner. Erlbaum. [YG-G]

Klahr, D. & McWhinney, B. (1998) Information processing. In: *Handbook of child psychology: Vol. 2. Cognition, perception, and language, 5th edition,* ed. D. Kuhn & R. S. Siegler. (W. Damon, Series Editor). Wiley. [DP-D]

Klayman, J. & Ha, Y.-W. (1987) Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review* 94:211–28. [MEG]

Koedinger, K. R., Anderson, J. R., Hadley, W. H. & Mark, M. A. (1997) Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education* 8:30–43. [MWA]

Koedinger, K. R. & MacLaren, B. A. (1997) Implicit strategies and errors in an improved model of early algebra problem solving. In: *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society,* ed. M. Shafto & P. Langley. Erlbaum. [MWA]

Koedinger, K. R., Nathan, M. J. & Alibali, M. W. (1997) Bridges to representational fluency: Grounding and abstraction in early algebra instruction. Proposal to the James S. McDonnell Foundation, Cognitive Studies for Educational Practice Program. [MWA]

Kolers, P. A. & Roedinger, H. L. (1984) Procedures of mind. *Journal of Verbal Learning and Verbal Behavior* 23:425–49. [SK]

Kosslyn, S. M. (1975) Information representation in visual images. *Cognitive Psychology* 7:341–70. [rJP]

Künne, W. (1995) Some varieties of thinking. Reflections on Meinong and Fodor. *Grazer Philosophische Studien* 50:365–97. [aZD]

LaBerge, D. (1997) Attention awareness and the triangular circuit. *Consciousness and Cognition* 6:149–81. [JT]

Laham, D. (1997) Latent semantic analysis approaches to categorization. In: *Proceedings of the 19th Annual Meeting of the Cognitive Science Society,* ed. M. G. Shafto & P. Langley. Erlbaum. [JRV]

Landauer, T. K. & Dumais, S. T. (1997) A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104:211–40. [JRV]

Lebiere, C. & Wallach, D. (1998) Implicit does not imply procedural: A declarative theory of sequence learning. In: *Proceedings of the Fifth Annual ACT-R Workshop,* ed. C. Lebiere. Carnegie Mellon University. [CL]

(in preparation) An integrated theory of sequence learning. [CL]

Lebiere, C., Wallach, D. & Taatgen, N. (1998) Implicit and explicit learning in ACT-R. In: *Proceedings of the Second European Conference on Cognitive Modelling,* ed. F. E. Ritter & R. M. Young. Nottingham University Press. [CL, NAT]

Leslie, A. M. (1987) Pretense and representation: The origins of "Theory of Mind." *Psychological Review* 94:412–26. [aZD, rJP]

(1994) Pretending and believing: Issues in the theory of ToMM. *Cognition* 50:211–38. [aZD]

Levine, J. M. & Moreland, R. L. (in press) Knowledge transmission in work groups: Helping newcomers to succeed. In: *Shared knowledge in organizations,* ed. L. Thompson, D. Messick & J. Levine. Erlbaum. [MEG]

Lewis, C. & Mitchell, P., eds. (1994) *Children's early understanding of mind: Origins and development.* Erlbaum. [DP-D]

Lewis, D. (1986) Causal explanation. In: *Philosophical papers, vol. 2,* ed. D. Lewis. Oxford University Press. [aZD]

Lewis, V. & Boucher, J. (1988) Spontaneous, instructed and elicited play in relatively able autistic children. *British Journal of Developmental Psychology* 6:325–39. [rJP]

Light, L. L. (1991) Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology* 42:333–76. [NWM]

Light, L. L. & Albertson, S. A. (1989) Direct and indirect tests of memory for category exemplars in young and older adults. *Psychology and Aging* 4:487–92. [NWM]

Lycan, W. (1988) *Judgement and justification.* Cambridge University Press. [SN]

Mackenzie, D. & Spinardi, G. (1995) Tacit knowledge, weapons design, and the uninvention of nuclear weapons. *American Journal of Sociology* 101(1):44–99. [MEG]

Mandler, J. M. (1992) How to build a baby: II. Conceptual primitives. *Psychological Review* 99:587–604. [DP-D]

(1998) Representation. In: *Handbook of child psychology: Vol. 2. Cognition, perception, and language, 5th edition,* ed. D. Kuhn & R. S. Siegler. (W. Damon, Series Editor). Wiley. [DP-D]

Manza, L. & Bornstein, R. F. (1995) Affective discrimination and the implicit learning process. *Consciousness and Cognition* 4:399–409. [RFB]

Manza, L. & Reber, A. S. (1997) Representation of tacit knowledge: Transfer across stimulus forms and modalities. In: *How implicit is implicit learning?,* ed. D. Berry. Oxford University Press. [aZD]

Marcel, A. J. (1983a) Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology* 15:197–237. [aZD, JT, CEW]

(1983b) Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology* 15:238–300. [aZD]

(1993) Slippage in the unity of consciousness. In: *Experimental and theoretical studies of consciousness: Ciba Foundation Symposium 174, ed.* G. R. Bock &. J. Marsh. Wiley. [aZD]

Mathis, W. D. & Mozer, M. C. (1995) On the computational utility of consciousness. In: *Advances in neural information processing systems 7,* ed. G. Tesauro, D. S. Touretzky & T. K. Leen. MIT Press. [DCN]

(1996) Conscious and unconscious perception: A computational theory. In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society,* ed. G. W. Cottrell. Erlbaum. [LJ]

Maunsell, J. H. R. & Newsome, W. T. (1987) Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience* 10:363–401. [CEW]

McCarthy, J. & Hayes, P. J. (1969) Some philosophical problems from the standpoint of artificial intelligence. In: *Machine intelligence, vol. 4,* ed. B. Mehler & D. Michie. Edinburgh University Press. [aZD]

McClelland, D. C., Koestner, R. & Weinberger, J. (1989) How do implicit and self-attributed motives differ? *Psychological Review* 96:690–702. [RFB]

McGeorge, P., Crawford, J. R. & Kelly, S. W. (1997) The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:239–45. [NAT]

Mercado, E., III, Murray, S. O., Uyeyama, R. K., Pack, A. A. & Herman, L. M. (1998) Memory for recent actions in the bottlenosed dolphin (*Tursiops truncatus*): Repetition of arbitrary behaviors using an abstract rule. *Animal Learning and Behavior* 26:210–18. [EM]

Mercado, E., III, Uyeyama, R. K., Pack, A. A. & Herman, L. M. (in press) Memory for action events in the bottlenosed dolphin. *Animal Cognition* 2:17–25. [EM]

Merikle, P. M. (1992) Perception without awareness: Critical issues. *American Psychologist* 47:792–95. [aZD]

Merriam-Webster Collegiate Dictionary (1994) *http://www.eb.com:180/cgi-bin/g?DocF=dict/ex/explicit.html* [GOv]

Merzenich, M. M. & deCharms, R. C. (1996) Neural representations, experience, and change. In: *Mind-brain continuum,* ed. R. Llinas & P. S. Churchland. MIT Press. [EM]

Millikan, R. G. (1984) *Language, thought, and other biological categories.* MIT Press. [aZD, NG]

Milner, D. A. & Goodale, M. A. (1995) Visual pathways to perception and action. In: *Progress in brain research, vol. 95,* ed. T. P. Hicks, S. Molotschnikoff & Y. Ono. Elsevier. [aZD, NG]

Moscovitch, M. (1995) Models of consciousness and memory. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. Bradford. [SK]

Mulligan, N. W. (1996) The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:1067–87. [NWM]

(1998) The role of attention during encoding on implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:27–47. [NWM]

Mulligan, N. W. & Hartman, M. (1996) Divided attention and indirect memory tests. *Memory and Cognition* 24:453–65. [NWM]

Munakata, Y., McClelland, J. L., Johnson, M. H. & Siegler, R. S. (1997) Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review* 104:686–713. [TR]

Newell, A. (1973) You can't play twenty questions with nature and win. In: *Visual information processing,* ed. W. C. Chase. Academic Press. [CL, NAT]

Nichols, S. & Stich, S. (1998) A cognitive theory of pretense. Unpublished manuscript, College of Charleston. [aZD, rJP]

Noelle, D. C. & Cottrell, G. W. (1994) Towards instructable connectionist systems. In: *Computational architecture integrating neural and symbolic processes,* ed. R. Sun & L. A. Bookman. Kluwer Academic. [DCN]

Norman, D. A. & Shallice, T. (1980) Attention to action: Willed and automatic control of behaviour. Center for Human Information Processing Technical Report No. 99. Reprinted in revised form in: *Consciousness and self-regulation, vol. 4,* ed. R. J. Davidson, G. E. Schwartz & D. Shapiro. Plenum, 1986. [aZD, rJP, JT, PDZ]

O'Brien, G. & Opie, J. (1999) A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* 22(1):127–96. [LJ, GO, JRV]

Olton, D. S., Markowsa, A. L., Pang, K., Golski, S., Voytko, M. L. & Gorman, L. K.

(1992) Comparative cognition and assessment of cognitive processes in animals. *Behavioural Pharmacology* 3:307–18. [EM]
Over, D. E. & Evans, J. St. B. T. (1997) Two cheers for deductive competence. *Current Psychology of Cognition* 16:255–78. [JSBTE]
Overskeid, G. (1994a) The intuitive mind. *Behavioral and Brain Sciences* 17:414. [GOv]
(1994b) Knowledge, consciousness, terminology, and therapy. *Scandinavian Journal of Behaviour Therapy* 23:65–72. [GOv]
(1995) Cognitivist or behaviourist - who can tell the difference? The case of implicit and explicit knowledge. *British Journal of Psychology* 86:517–22. [GOv]
(1999) Forklaring, lovmessighet og det selvfylgelige i psykologisk forskning og praksis. [Explanation, lawfulness, and the self-evident in psychological research and practice]. *Tidsskrift for Norsk Psykologforening* 38:42–44. [GOv]
Paillard, J., Michel, F. & Stelmach, G. (1983) Localization without content: A tactile analogue of "blindsight." *Archives of Neurology* 40:548–51. [aZD]
Patton, M. J. & Jackson, A. P. (1991) Theory and meaning in counseling research: Comment on Strong (1991). *Journal of Counseling Psychology* 38:214–16. [GOv]
Peacocke, C. (1991) *A study of concepts.* MIT Press. [rJP]
Perner, J. (1990) Experiential awareness and children's episodic memory. In: *Interactions among aptitudes, strategies, and knowledge in cognitive performance,* ed. W. Schneider & F. E. Weinert. Springer Verlag. [aZD]
(1991) *Understanding the representational mind.* MIT Press/A Bradford Book. [JB, aZD, rJP]
(1995) The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition* 57:241–69. [aZD]
(1998) The meta-intentional nature of executive functions and theory of mind. In: *Language and thought,* ed. P. Carruthers & J. Boucher. Cambridge University Press. [aZD, rJP]
Perner, J. & Clements, W. A. (1999) From an implicit to an explicit theory of mind. In: *Beyond dissociation: Interaction between dissociated implicit and explicit processing,* ed. Y. Rossetti & A. Revonsuo. John Benjamins. [aZD, NG]
Perner, J. & Lang, B. (1999) What accounts for the developmental relationship between theory of mind and executive function? Paper presented in the symposium "Executive function and theory of mind" at the Biennial Meeting of the Society for Research in Child Development (SRCD), Albuquerque, New Mexico, April 15–18, 1999. [rJP]
(in press) Theory of mind and executive function: Is there a developmental relationship? In: *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience,* ed. S. Baron-Cohen, H. Tager-Flusberg & D. Cohen. Oxford University Press. [rJP]
Perner, J., Leekam, S. R. & Wimmer, H. (1987) Three-year olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology* 5:125–37. [aZD]
Perner, J., Stummer, S. & Lang, B. (1999) Executive functions and theory of mind: Cognitive complexity of functional dependence? In: *Developing theories of intention: Social understanding and self-control,* ed. P. D. Zelazo, J. W. Astington & D. R. Olson. Erlbaum. [rJP]
Perruchet, P. & Amorin, P. (1992) Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:785–800. [CL]
Perruchet, P. & Gallego, J. (1997) A subjective unit formation account of implicit learning. In: *How implicit is implicit learning?,* ed. D. Berry. Oxford University Press. [aZD]
Perruchet, P. & Pacteau, C. (1990) Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General* 119:264–75. [P-JM, GO]
Perruchet, P., Vintner, A. & Gallego, J. (1997) Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin and Review* 4:43–48. [P-JM]
Perry, J. (1986) Thought without representation. *Supplementary Proceedings of the Aristotelian Society* 60:137–66. [aZD]
Perry, M., Church, R. B. & Goldin-Meadow, S. (1988) Transitional knowledge in the acquisition of concepts. *Cognitive Development* 3:359–400. [MWA, SG-M]
Piaget, J. (1923) La psychologie des valeurs religieuses. In: *Sainte-Croix 1922,* ed. Association Chrétienne d'Etudiants de la Suisse Romande, 38–82. [LS]
(1945) *Play, dreams, and imitation in childhood.* W. W. Norton. [rJP]
(1954) *The construction of reality in the child.* Basic Books. [SG-M]
(1976) *The grasp of consciousness: Action and concept in the young child.* Harvard University Press. [BDH]
(1986) Essay on necessity. *Human Development* 29:301–14. [LS]
(1995) *Sociological studies.* Routledge. [LS]
Piaget, J. & Inhelder, B. (1941/1974) *The child's construction of quantities: Conservation and atomism,* trans. A. J. Pomerans. Basic Books. [aZD]
Plato (1986) *Meno,* trans. R. W. Sharples. Aris & Phillips. [BDH]
Pöppel, E. (1988) *Mindworks: Time and conscious experience.* Harcourt Brace Jovanovich. [RAC]
Pöppel, E., Held, R & Frost, D. (1973) Residual visual function after brain wounds involving the central visual pathways in man. *Nature* 243:295–96. [aZD]
Poulin-Dubois, D. (1999) Infants' distinction between animate and inanimate objects: The origins of naive psychology. In: *Early social cognition,* ed. P. Rochat. Erlbaum. [DP-D]
Poulin-Dubois, D., Frank, I., Graham, S. A. & Elkin, A. (1999) The role of shape similarity in toddlers' lexical extensions. *British Journal of Developmental Psychology* 17:21–36. [DP-D]
Prinz, W. (1990) A common coding approach to perception and action. In: *Relationships between perception and action: Current approaches,* ed. O. Neumann & W. Prinz. Springer-Verlag. [rJP]
Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* 80:1–24. [rJP]
(1978) When is attribution of beliefs justified? *Behavioral and Brain Sciences* 1:592–93. [aZD]
(1984) *Computation and cognition.* MIT Press. [GO]
Quine, W. V. O. (1951) Two dogmas of empiricism. *Philosophical Review* 60:20–43. [rJP]
(1956) Quantifiers and propositional attitudes. *Journal of Philosophy* 53:177–87. [AB]
Rajaram, S. (1996) Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:365–77. [NWM]
Rakison, D. H. & Butterworth, G. E. (1998) Infants' use of object parts in early categorization. *Developmental Psychology* 34:49–62. [DP-D]
Reason, J. T. & Mycielska, K. (1982) *Absent minded? The psychology of mental lapses and everyday errors.* Prentice Hall. [aZD]
Reber, A. S. (1967) Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour* 6:855–63. [aZD, GO, JRV]
(1969) Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology* 81:115–19. [JRV]
(1976) Implicit learning of artificial grammars: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory* 2:88–94. [JRV]
(1989) Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General* 118:219–35. [aZD, P-JM, GO]
(1993) *Implicit learning and tacit knowledge.* Oxford University Press. [aZD, JSBTE]
Reber, A. S., Kassin, S. M., Lewis, S. & Cantor, G. (1980) On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory* 6(5):492–502. [DCN]
Reder, L. M. (1988) Strategic control of retrieval strategies. *The Psychology of Learning and Motivation* 22:227–59. [PAB]
Reingold, E. M. & Merikle, P.M. (1988) Using direct and indirect measures to study perception without awareness. *Perception and Psychophysics* 44:563–75. [aZD]
Reingold, E.M. & Merikle, P.M. (1993) Theory and measurement in the study of unconscious processes. In: *Consciousness,* ed. M. Davies & G. W. Humphreys. Blackwell. [aZD]
Richardson-Klavehn, A. & Bjork, R. A. (1988) Measures of memory. *Annual Review of Psychology* 39:475–543. [aZD]
Richardson-Klavehn, A., Gardiner,J. M. & Java, R. I. (1994) Involuntary conscious memory and the method of opposition. *Memory* 2:1–29. [aZD, SK]
(1996) Memory: Task dissociations, process dissociations, and dissociations of consciousness. In: *Implicit cognition,* ed. G. Underwood. Oxford University Press. [aZD]
Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3:131–41. [NG]
Roberts, P. L. & McLeod, C. (1995) Representational consequences of two modes of learning. *Quarterly Journal of Experimental Psychology* 48A:296–319. [aZD, rJP]
Roediger, H. L. (1990) Implicit memory: Retention without remembering. *American Psychologist* 45:1043–56. [SK, NAT]
Roediger, H. L. & Blaxton, T. A. (1987) Retrieval modes produce dissociations in memory for surface information. In: *Memory and learning: The Ebbinghaus Centennial Conference,* ed. D. S. Gorfein & R. R. Hoffman. Erlbaum. [SK]
Roediger, H. L. & McDermott, K. B. (1993) Implicit memory in normal human subjects. In: *Handbook of neuropsychology, vol. 8,* ed. F. Boller & J. Grafman. Elsevier. [YG-G, NWM]
(1996) Implicit memory tests measure incidental retrieval. Paper presented at the XXVI International Congress of Psychology, Montreal, August 1996. [aZD]
Roitblat, H. L. (1987) *Introduction to comparative cognition.* Freeman. [EM]

Rosch, E. & Mervis, C. B. (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7:573–605. [JAH]
Rosenthal, D. M. (1986) Two concepts of consciousness. *Philosophical Studies* 49:329–59. [aZD, GO]
Rossetti, Y. (1998) Implicit short-lived motor representation of space in brain-damaged and healthy subjects. *Consciousness and Cognition* 7:520–58. [aZD, NG]
Ruffman, T. (1996) Do children understand the mind by means of a theory or simulation?: Evidence from their understanding of inference. *Mind and Language* 11:388–414. [TR]
Ruffman, T., Clements, W. A., Import, A. & Connolly, D. (1998) Does eye direction indicate implicit sensitivity to false belief? Unpublished manuscript, University of Sussex. [rJP, TR]
Russell, B. (1961) Knowledge by acquaintance and knowledge by description. In: *The basic writings of Bertrand Russell: 1903–1959,* ed. R. E. Egner & L. E. Denonn. Simon & Schuster. (Original work published in 1912). [GOv]
(1964) *The principles of mathematics,* 2nd edition. Norton. [LS]
(1991) On propositions: What they are and what they mean. *Proceedings of the Aristotelian Society* 2:1–43. [aZD]
Russell, J., Mauthner, N., Sharpe, S. & Tidswell, T. (1991) The "windows task" as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology* 9:331–49. [rJP]
Sachs, J. S. (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2:437–42. [JT]
Salmon, W. C. (1984) *Scientific explanation and the causal structure of the world.* Princeton University Press. [aZD]
Schacter, D. L. (1987) Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:501–18. [RFB, aZD, NWM]
Schacter, D. L., Bowers, J. & Booker, J. (1989) Intention, awareness, and implicit memory: The retrieval intentionality criterion. In: *Implicit memory: Theoretical issues,* ed. S. Lewandowsky, J. C. Dunn & K. Kirsner. Erlbaum. [aZD, SK]
Schacter, D. L. & Tulving, E. (1994) *Memory systems.* MIT Press/Bradford Books. [YG-G]
Schiffer, S. (1987) *Remnants of meaning.* MIT Press. [AB]
Schneider, K. (1959) *Clinical psychopathology.* Grune & Stratton. [NG]
Schwarz, B. L., Rosse, R. B. & Deutsch, S. I. (1993) Limits of the processing view in accounting for dissociations among memory measures in a clinical population. *Memory and Cognition* 21:63–72. [NWM]
Searle, J. (1983) *Intentionality.* Cambridge University Press. [RAC, aZD]
Shallice, T. (1988) Specialisation within the semantic system. Special Issue: The cognitive neuropsychology of visual and semantic processing of concepts. *Cognitive Neuropsychology* 5:133–42. [aZD]
Shanks, D. R. (1995) *The psychology of associative learning.* Cambridge University Press. [aZD]
Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–448. [aZD, P-JM, DCN, GOv]
Shimamura, A. P. (1986) Priming effects in amnesia: Evidence for a dissociable memory function. *Quarterly Journal of Experimental Psychology* 38A:619–44. [NWM]
(1993) Neuropsychological analyses of implicit memory: History, methodology, and theoretical interpretations. In: *Implicit memory: New directions in cognition, development, and neuropsychology,* ed. P. Graf & M. E. J. Masson. Erlbaum. [NWM]
Singley, M. R. & J. R. Anderson (1989) *The transfer of cognitive skill.* Harvard University Press. [MEG]
Skinner, B. F. (1950) Are theories of learning necessary? *Psychological Review* 7:193- 216. [GOv]
Sloman, S. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [aZD, JSBTE, CL]
Smedslund, J. (1998) Hvorfor klinisk forskning og praksis ikke går hånd i hånd. [Why clinical research and practice do not go hand in hand]. *Tidsskrift for Norsk Psykologforening* 35:1090–95. [GOv]
Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R. & Erb, L. (1995) The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General* 124:391–408. [EM]
Smith, L. (1992) Judgments and justifications: Criteria for the attribution of children's knowledge in Piagetian research. *British Journal of Developmental Psychology* 10:1–23. [LS]
(1993) *Necessary knowledge: Piagetian perspectives on constructivism.* Erlbaum. [LS]
(1998) On the development of mental representation. *Developmental Review* 18:202–27. [LS]
(1999a) Epistemological principles for developmental psychology in Frege and Piaget. *New Ideas in Psychology.* (in press). [LS]
(1999b) Necessary knowledge in number conservation. *Developmental Science* 2(1):23–27. [LS]
(1999c) What Piaget learned from Frege. *Developmental Review* 19:133–53. [LS]
Smith, N. & Tsimpli, I.-A. (1995) *The mind of a savant: Language-learning and modularity.* Blackwell. [aZD]
Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. (1992) Origins of knowledge. *Psychological Review* 99:605–32. [SG-M]
Spelke, E. S. & Kestenbaum, R. (1986) Les origines du concept d'objet. *Psychologie Francaise* 31:67–72. [rJP]
Spelke, E. S., Phillips, A. & Woodward, A. L. (1995) Infant's knowledge of object motion and human action. In: *Causal cognition: A multidisciplinary debate,* ed. D. Sperber, D. Premack & A. J. Premack. Oxford University Press. [rJP]
Sperber, D. (1996) *Explaining culture: A naturalistic approach.* Blackwell. [aZD]
(1997) Intuitive and reflective beliefs. *Mind and Language* 12(1):67–83. [aZD, rJP]
Squire, L. R. (1987) *Memory and brain.* Oxford University Press. [MAS]
(1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99(2):195–231. [aZD, EM, SN, rJP]
Squire, L. R. & Knowlton, B. J. (1995) Memory, hippocampus, and brain systems. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [NAT]
Srinivas, K. & Roediger, H. L. (1990) Classifying implicit memory tests: Category association and anagram solution. *Journal of Memory and Language* 29:389–412. [NWM]
Stadler, M. A. & Frensch, P. A., eds. (1998) *Handbook of implicit learning.* Sage. [aZD]
Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning.* Erlbaum. [JSBTE]
Stanovich, K. E. & West, R. F. (1998) Cognitive ability and variation in the selection task. *Thinking and Reasoning* 4:193–288. [JSBTE]
Strawson, P. F. (1959) *Individuals.* Methuen. [AB, aZD]
Taatgen, N. A. (1999) Learning without limits: From problem solving towards a unified theory of learning. Unpublished thesis, University of Groningen, The Netherlands. http://tcw2.ppsw.rug.nl/~niels/thesis. [NAT]
Tulving, E. (1985) Memory and consciousness. *Canadian Psychology* 26:1–12. [aZD, rJP]
Tulving, E., Schacter, D. L. & Stark, H. A. (1982) Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:336–42. [YG-G, rJP, NAT]
Tweney, R. D. (1985) Faraday's discovery of induction: A cognitive approach. In: *Faraday rediscovered: Essays on the life and work of Michael Faraday: 1791–1867,* ed. D. Gooding & F. James. Stockton Press. [MEG]
Tye, M. (1995) *Ten problems of consciousness: A representational theory of the phenomenal mind.* MIT Press. [aZD]
Tzelgov, J. (1997) Specifying the relations between automaticity and consciousness: A theoretical note. *Consciousness and Cognition* 6:441–51. [JT]
Tzelgov, J., Porat, Z. & Henik, A. (1997) Automaticity and consciousness: Is perceiving the word necessary for reading it? *American Journal of Psychology* 110:429–48. [rJP, JT]
Ungerleider, L. & Mishkin, M. (1982) Two cortical visual systems. In: *Analysis of motor behavior,* ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield. MIT Press. [aZD, MAS]
Vallacher, R. & Wegner, D. (1987) What do people think they are doing? Action identification and human behavior. *Psychological Review* 94:3–15. [JT]
Vaughan, D. (1996) *The Challenger launch decision.* The University of Chicago Press. [MEG]
Vokey, J. R. & Brooks, L. R. (1992) Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:328–44. [P-JM, JRV]
(1994) Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:1504–10. [JRV]
Vokey, J. R. & Read, J. D. (1995) Memorability, familiarity, and categorical structure in the recognition of faces. In: *Cognitive and computational aspects of face recognition,* ed. T. Valentine. Routledge. [JRV]
Wallach, D. & Lebiere, C. (1998) Modellierung von Wissenserwerbsprozessen bei der Systemregelung. [Modeling knowledge acquisition in system control]. In: *Intelligente Informationsverarbeitung* [*Intelligent information processing*], ed. W. Krause & U. Kottkamp. Deutscher Universitätsverlag. [CL]
Warrington, E. K. & Weiskrantz, L. (1970) Amnesia: Consolidation or retrieval? *Nature* 228:628–30. [NAT]
Weiskrantz, L. (1988) Some contributions of neuropsychology of vision and memory to the problem of consciousness. In: *Consciousness in contemporary science,* ed. A. J. Marcel & E. Bisiach. Clarendon Press. [aZD]
Weiskrantz, L., Warrington, E. K., Sanders, M. D. & Marshall, J. (1974) Visual capacity in hemianopic field following a restricted occipital ablation. *Brain* 97:709–28. [aZD]

Weldon, M. S. (1991) Mechanisms underlying priming on perceptual tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:526–41. [YG-G]

Weldon, M. S. & Coyote, K. C. (1996) Failure to find the picture superiority effect in implicit conceptual memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:670–86. [NWM]

Wellman, H. M. & Woolley, J. D. (1990) From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition* 35:245–75. [DP-D]

Whiten, A. & Byrne, R. W., eds. (1997) *Machiavellian intelligence II: Extensions and evaluations.* Cambridge University Press. [AC-M]

Whittlesea, B. W. A. & Dorken, M. D. (1993) Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General* 122:227–48. [aZD, P-JM, JRV

Willingham, D. B., Nissen, M. J. & Bullemar, P. (1989) On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:1047–60. [CL]

Winograd, T. (1975) Frame representations and the declarative-procedural controversy. In: *Representation and understanding: Studies in cognitive science,* ed D. G. Bobrow & A. Collins. Academic Press. [aZD]

Wittgenstein, L. (1958) *Philosophical investigations,* 3rd edition. Blackwell. [GOv]

Wong, E. & Mack, A. (1981) Saccadic programming and perceived location. *Acta Psychologica* 48:123–31. [aZD]

Wright, R. L. & Burton, M. A. (1995) Implicit learning of an invariant: Just say no. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 48A:783–96. [JRV]

Xitco, M. J., Jr. (1988) *Mimicry of modeled behaviors by a bottlenosed dolphin.* M. A. thesis, University of Hawaii, Honolulu. [EM]

Xu, F. & Carey, S. (1996) Infants' metaphysics: The case of numerical identity. *Cognitive Psychology* 30:111–53. [rJP]

Zajonc, R. B. (1968) Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monographs* 9(2, pt. 2):1–27. [aZD]

Zeki, S. M. (1978) Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *Journal of Physiology* 277:90. [CEW]

Zelazo, P. D. (1996) Towards a characterization of minimal consciousness. *New Ideas in Psychology* 14:63–80. [PDZ]

Zelazo, P. D., Astington, J. W. & Olson, D. R. (1999) *Developing theories of intention: Social understanding and self-control.* Erlbaum. [PDZ]

Zelazo, P. D. & Frye, D. (1997) Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In: *Language structure, discourse, and the access to consciousness,* ed. M. Stamenov. John Benjamins. [PDZ]

Zelazo, P. D., Frye, D. & Rapus, T. (1996) An age-related dissociation between knowing rules and using them. *Cognitive Development* 11:37–63. [PDZ]

Zelazo, P. D., Reznick, J. S. & Pinon, D. E. (1995) Response control and the execution of verbal rules. *Developmental Psychology* 31:508–17. [aZD]

Zelazo, P. R. & Zelazo, P. D. (1998) The emergence of consciousness. In: *Consciousness: At the frontiers of neuroscience. Advances in neurology, vol. 77,* ed. H. H. Jasper, L. Descarries, V. F. Castellucci & S. Rossignol. Lippincott- Raven Press. [PDZ]

Zola-Morgan, S. & Squire, L. (1990) The neuropsychology of memory: Parallel findings in humans and nonhuman primates. In: *The development and neural bases of higher cognitive functions,* ed. A. Diamond. New York Academy of Sciences. [SN]

https://doi.org/10.1017/S0140525X9957218X Published online by Cambridge University Press