Zoltán Dienes

Assumptions of Subjective Measures of Unconscious Mental States

Higher Order Thoughts and Bias

Abstract: *This paper considers two subjective measures of the existence of unconscious mental states — the guessing criterion, and the zero correlation criterion — and considers the assumptions underlying their application in experimental paradigms. Using higher order thought theory (Rosenthal, e.g. 1986; 1995) the impact of different types of biases on the zero correlation and guessing criteria are considered. It is argued that subjective measures of consciousness can be biased in various specified ways, some of which involve the relation between first order states and second order thoughts, and hence are not errors in measurement of the conscious status of mental states; but other sorts of biases are measurement errors, involving the relation between higher order thoughts and their expression. Nonetheless, it is argued this type of bias does not preclude subjective measures — both the guessing criterion and the zero correlation criterion — as being amongst the most appropriate and useful tools for measuring the conscious status of mental states.*

Keywords: higher order thoughts, consciousness, implicit learning, subliminal perception, subjective threshold, confidence, artificial grammar learning

Introduction

The question of how we should determine the conscious or unconscious status of mental states has a long history in experimental psychology, going back to the beginning of the discipline (e.g. Peirce & Jastrow, 1884). Measures of the conscious status of mental states can be divided into objective measures (based on better than chance discrimination of stimuli) and subjective measures (based on subjects' reports of the mental states by which they may have beheld any

Correspondence:
Zoltán Dienes, Department of Psychology, Sussex University, Brighton, BN1 9QG, UK.
E-mail: dienes@sussex.ac.uk

Journal of Consciousness Studies, **11**, No. 9, 2004, pp. 25–45

stimuli). For example, in the subliminal perception case, if the subject can objectively discriminate amongst possible choices for a just displayed object, the perception is regarded as conscious by objective measures. But the perception is only regarded as conscious by subjective measures if the person believes they based their choices on a mental state that provides knowledge, e.g. one of seeing (i.e., if they believe they based their choice on a visual experience, see e.g. Merikle, 1992).

There are two basic subjective measures of conscious states in the perception and learning literatures: the guessing criterion and the zero correlation criterion (Dienes & Berry, 1997). According to the guessing criterion, subjects have unconscious knowledge if they perform above baseline while claiming to be purely guessing. Although the guessing criterion has been around for a long time (e.g. Peirce & Jastrow, 1884), its persuasive introduction into the modern experimental literature was accomplished by Cheesman and Merikle (1984; 1986; see also Weiskrantz, 1986; 1997; Merikle, 1992). Cheesman and Merikle presented subjects with one of four colour words followed after some Stimulus Onset Asynchrony (SOA) by a mask. After each word subjects reported which word they thought had been presented. After each block of trials, subjects estimated their percent correct classification, where 25% corresponded to chance and hence indicated subjects believed they were basing their responses on no relevant information whatsoever (i.e. not on seeing but just on guessing), and 100% corresponded to complete confidence. After each block, the SOA was reduced until subjects claimed they were performing at chance. They found that at this subjective threshold, subjects were still objectively discriminating at above chance (66% correct, compared to 25% baseline); i.e. despite the fact that the subjects were not consciously seeing, they were still picking up and using information. Dienes *et al*. (1995; also Dienes & Berry, 1997; Dienes & Altmann, 1997; Dienes & Perner, 2003; Tunney & Shanks, in press) applied the same criterion to the knowledge acquired about an artificial grammar. Subjects were initially asked to look at strings of letters which, unbeknownst to subjects, were generated by an artificial grammar (rules constraining the possible sequence of letters). After some minutes exposure to a set of strings, subjects were told about the existence of the grammar and asked to classify new strings. Subjects could do this at an above chance rate (e.g. 60%), even when they claimed they were guessing; i.e. subjects were using knowledge they did not know they had, satisfying the guessing criterion of unconscious knowledge.

The zero correlation criterion looks at the within-subject relationship between confidence and accuracy. Knowledge is said to be unconscious if there is no relation between confidence and accuracy. In an artificial grammar learning task, Chan (1992) asked for confidence ratings after each classification decision. He found that subjects were no more confident in correct rather than incorrect decisions, and took this to be an indication of unconscious knowledge. Dienes *et al.* (1995), Dienes and Altmann (1997), Allwood *et al*. (2000), Channon *et al.* (2002), Tunney and Altmann (2001) and Dienes and Perner (2003) replicated these results, finding some conditions under which there was no within-subject

relationship between confidence and accuracy. We argued this indicated subjects could not discriminate between mental states providing knowledge and e.g. those just corresponding to guessing; hence, the mental states were unconscious (see also Kelley *et al*., 2001, and Newell & Bright, 2002, who used the same lack of relationship between confidence and accuracy to argue for the use of unconscious knowledge in other learning paradigms). Kolb and Braun (1995) and Kunimoto *et al*. (2001) applied a similar methodology to perception; they found conditions under which confidence was not related to accuracy, and argued this demonstrated the existence of unconscious perception. (All the above papers also found conditions in which there was conscious knowledge by the same criterion; see also Redington *et al.*, 1996; Tunney & Shanks, in press.)

According to the logic of subjective measures of consciousness, mental states are shown as being conscious if the subject is conscious of properties of mental states (rather than properties of merely the stimuli the states are about). Thus, accepting that subjective measures have an advantage over objective measures in determining whether mental states are conscious or not, prima facie involves accepting some sort of higher order theory of consciousness (i.e. theories in which it is being conscious of a mental state rather than of the stimuli the state is about that is crucial for a mental state to be a conscious mental state). Initially, I will describe higher order theories of consciousness, focusing on Rosenthal's theory (e.g. 1986; 2000a,b,c; forthcoming). This will provide a conceptual framework for considering the assumptions underlying the use of the guessing criterion and the zero correlation criterion in detecting unconscious mental states. One common criticism of subjective measures is that they are biased (e.g. Eriksen, 1960; Reingold & Merikle, 1993) and this is sometimes taken to rule out their usefulness (e.g. Dulany, 1997). Using higher order thought theory, I will discuss a number of ways in which subjective measures can be biased. Next I will indicate the evidence regarding whether subjective measures are actually biased. Finally, I argue that potential or actual bias does not rule out subjective measures as amongst the most appropriate tools for detecting unconscious mental states.

Initially, higher order thought theory will be described, and then the arguments listed above will be made in turn.

Higher Order Thought (HOT) Theory

What would make the perception of the word in front of you being 'butter' a conscious perception? In general, under what conditions is a mental state (sensation, thought, desire, etc.) a conscious mental state? An answer to this question is provided by higher order theories of consciousness (e.g. Armstrong, 1980; Rosenthal, 1986; 2000a,b,c; Carruthers, 1992; 2000; see Block, 2001, and Rowlands, 2001, for recent criticism, and Gennaro's edited volume, 2004, for arguments for and against). I will focus on Rosenthal's higher order thought (HOT) theory, a philosophical theory of consciousness that is appealing for its simplicity and elegance. In order to have an account of a mental state, like a thought, being conscious, we need to consider the logical possibility of thoughts

being unconscious, so we can consider what would make the mental state conscious independently of simply being a mental state. That is, to say that someone is thinking, we should not presume that they must be consciously thinking; they could be unconsciously thinking. With that proviso in mind, we can consider how we become conscious of events and things. In general, I can be conscious of things in two ways; by perception and by thinking. I can be conscious of a problem by thinking about a problem; I can be conscious of you by seeing you or just by thinking about you being there. If we flash a person either the word 'butter' or the word 'grass', and they can later make a forced choice discrimination above chance about the identity of the word, we can say he is conscious of the word because he saw the word. (Hence, the apparent appeal of objective tests; they reflect whether the subject is conscious of the stimulus.) But by 'conscious of the word' we do not necessarily mean consciously aware of the word or that he beheld the word with a conscious mental state. In a sense he is conscious of the word; but the seeing itself need not be a conscious mental state. Rosenthal appeals to a common (though not universal: e.g. Block, 2001) intuition that for a mental state to be a conscious mental state, we should be conscious of the mental state. We could not claim that a person has a conscious mental state, and also claim that the person is not conscious of being in the mental state. According to Rosenthal, the relevant way of being conscious of the mental state is to have a thought about the mental state. For example, if the mental state is seeing that the word is butter, one becomes conscious of the mental state by thinking 'I *see* that the word is butter'; because the state of affairs of the word being butter is now beheld with a conscious mental state, the person is consciously aware of the word being butter. In general, according to the theory, I am consciously aware of a fact in virtue of entertaining a second order thought that represents me as being in the first order mental state (in the example, the first order mental state is seeing that the word is butter; the second order thought is representing that I am seeing that the word is butter).

The second order thought does not make itself, the second order thought, conscious, it just makes the first order thought that it is about conscious. The second-order thought 'I see that the word is butter' only makes one consciously aware of the word being butter, not the fact that one sees that the word is butter. The higher order thought theory is not just a theory of 'reflective consciousness'; it is a theory of how any mental state can be a conscious mental state. To be consciously aware that one knows by seeing, there needs to be a third order thought that makes the second order thought conscious. Reflection involves at least third order thoughts. The highest order thought the person has about a mental state is always itself unconscious on the theory, because there is not a yet higher order thought rendering it conscious.

We now use the higher order thought theory to consider the issues outlined above regarding the use of subjective measures of conscious awareness, i.e. the guessing criterion and the zero correlation criterion. The arguments do not depend on the full strong version of the higher order thought theory. According to Rosenthal's theory, the lower order state is conscious in virtue of the higher

order thought. One might postulate a less direct relationship between higher order thoughts and conscious awareness such that while every conscious mental state has an associated higher order thought, it is some third factor that causes both the first order state to be conscious and the higher order thought to occur (e.g. the third factor might be the first order state being attended to in a mechanism that is highly inferentially promiscuous, and thus always, as it happens, generates HOTs). This weaker theory is also sufficient to motivate the arguments that follow. Another type of weaker theory is a potentialist one, like that of Carruthers (2000): A mental state is conscious if it has the potential to elicit appropriate HOTs, whether or not those HOTs actually occur. In the experiments discussed in this paper, HOTs are always probed for, so the distinction between potential or occurrent HOTs being necessary is not important for most of the arguments that follow (a proviso is noted below): The probe will ensure any potential is actualised, so the presence of an actual HOT would be a good measure of whether there existed a potential HOT.[1]

A yet weaker theory is that while conscious first order states typically are associated with HOTs, they sometimes are not so associated (e.g. Block, 2001). A person who subscribed to this approach would not find the conclusions that follow convincing. If this view distinguished two different 'types' of conscious states, i.e. (1) conscious mental states that do not generate HOTs (so a person could sincerely deny they were in such a state although it was a conscious state) and (2) conscious states that are associated with HOTs (according to Block, 2001, the latter are 'reflexive' conscious states, and the former are non-reflexive conscious states), then the arguments that follow can be read as applying only to the second type of conscious state; the first type will be classified together with unconscious mental states in what follows.

Another type of higher order theory is higher order perception theory (e.g. Armstrong, 1980), according to which we become conscious of mental states by perceiving them (rather than thinking about them). In that case, subjective measures rely on expressing conceptual content in higher order perceptions in a confidence rating, and the arguments that follow can still be motivated (with references to higher order thoughts below being replaced with higher order perceptions).

Establishing that a subject can discriminate what stimulus has been presented (objective tests of conscious awareness) only establishes the presence of first order states: Representations about the world. According to higher order theories, such first order representations are not in themselves sufficient to provide conscious mental states. Blindsight patients (Weiskrantz, 1997) can discriminate stimuli in their 'blind' field at accuracies that can be close to 100%. We do not regard such patients as being consciously aware of the stimuli however, because they claim not to see anything at all. In other words, it is the absence of second order representations by the person THAT they are seeing that leads to the intuition that the seeing is unconscious. The mere demonstrable presence of first

[1] On Carruthers' (2000) theory, all HOTs are completely accurate about their first order states. So the problem of inaccurate HOTs, discussed in detail below, simply does not arise.

order visual representations, no matter how stable and enduring they are, does not in itself lead us to regard the seeing as conscious. It is subjective rather than objective measures that get to the heart of what it is for a mental state to be conscious. After all, consciousness is a subjective state.

Now we will consider some assumptions that underlie the use of subjective measures in psychology experiments. First we will consider biases that can arise in the formation and expression of second order thoughts in the absence of third order thoughts. Then we will consider the implication of the subjects' third order thoughts in generating bias.

Biases In the Formation and Expression of Second Order Thoughts

Higher order thoughts in general need not be all or none. There are three senses of not being all or none that I now specify. First, second order thoughts can represent thoughts as having graded attitudes, not purely knowing or purely guessing. There is a potential problem in translating a graded HOT into a verbal measure, a confidence rating (this process corresponds to the final stage of the Koriat *et al.*, 1980, model of confidence ratings). Translation is less problematic for the extreme attitudes of pure guess or complete certainty. If the subject understands what guessing means, it is conceptually straightforward to translate a *precisely-defined* graded HOT into a 'pure guess' response: Either the HOT represents the first order attitude as being to some degree knowing or not. However, subjects may choose for the 'guess' response category to include some range of attitudes, thereby introducing error in their reports. This may be the sort of error that Eriksen (1960), Holender (1986), Shanks & St John (1994), and Kunimoto *et al.* (2001) were referring to in describing reports of guessing as being possibly biased. This error is likely to be larger for coarse grained rather than fine grained confidence scales (but contrast Sahraie *et al.*, 1998, who found no effect of changing the coarseness of the scale with a blindsight patient) and will also depend on how subjects are instructed (about the meaning of 'guessing').

In the middle of the scale it will be yet more difficult for a subject to know how to translate different represented gradations of attitude into specific numerical ratings. In terms of the zero correlation criterion, binary measures of confidence (like that used by Kunimoto *et al.*, 2001, and Tunney & Shanks, in press) or ordinal measures of the confidence-accuracy relationship (e.g. 'Kendall's Tau', Kolb & Braun, 1995, the 'Goodman-Kruskal Gamma', Twyman, 2001, or 'Spearman's Rho', Kuhn & Dienes, 2003) need assume only that the subject respects ordinal relations between attitudes. The exact numbers, above guessing, do not matter.[2] However, variability in how represented attitudes are translated into verbal reports will introduce error into the zero correlation criterion: If the same represented attitude is sometimes referred to as high confidence and

[2] From the point of view of higher order thought theory, the exact number assigned an attitude above guessing is not relevant for determining the conscious status of the mental state. Thus, whether subjects are well calibrated or not, as defined in the confidence literature, is not important for assessing the conscious status of mental states. However, distinguishing guessing from some degree of knowing is important.

sometimes as low confidence, the zero correlation criterion will over-estimate the amount of unconscious knowledge.[3] Variability in how represented attitudes are translated into verbal report can reduce a positive confidence-accuracy relationship, but not reduce it to zero. However, given there is always a finite degree of experimental sensitivity, it can reduce the relationship to apparently zero.

Second, Rosenthal (1995) pointed out that there may be conflicting higher order thoughts about the same first order state. A person may simultaneously entertain two contradictory second order thoughts, e.g. one about guessing and one about knowing (knowing to some degree). An experimental probe may stop one of the second order thoughts from existing by eliciting the other. If this assumption is false, the guessing and zero correlation criteria may over-estimate the amount of unconscious knowledge. For the guessing criterion, the subject may report one HOT that is about guessing, while there exists another HOT that is about knowing and which renders the first order state conscious. For the zero correlation criterion, the report of the inaccurate HOT on some proportion of the time will introduce noise which flattens the measured accuracy confidence relationship, again making mental states seem unconscious when they are really conscious (given the finite degree of experimental sensitivity in measuring any particular confidence accuracy relationship).

Third, sometimes there may be no fact of the matter as to exactly which second order thought the subject has (guessing or some degree of knowing), or whether a second order thought has occurred at all (cf Dennett & Kinsbourne, 1992). At that moment, according to HOT theory, there is no fact of the matter as to whether the first order state is conscious. The first order state is more or less conscious (and correspondingly less or more unconscious) (Kosko, 1994). For example, if the forming of the HOT is interrupted at a certain time by a subsequent stimulus, there will be some indeterminacy about when the HOT has actually fully formed (150 or 190ms or 201 ms?). Probing for HOTs under leisurely conditions normally results in well-formed HOTs that have content that is as determinate as the person's concepts about mental states: Eliciting a well specified confidence rating forces HOTs with clear concepts, e.g. 'pure guess' etc. When subjects are in extreme conditions involving knowledge states formed on the basis of tiny transient amounts of evidence (as may happen in the laboratory), the process for forming HOTS may be on a knife edge. In the latter case, leisurely conditions for forming HOTs is likely to result in well formed HOTs, but whether the HOT is one of guessing or not could vary from trial to trial (and hence the conscious status of the first order state will vary from trial to trial).

[3] The extent to which this translation issue is regarded as a problem depends on the relation between thought and language (see e.g. Carruthers, 2002, for discussion of different theories). For example, if the content of a higher order thought is constituted by its linguistic logical form, then there is no problem of translating the thought into language: The thought is already in language form. However, the subject would have to map the words used internally into the words specified by the experimenter for verbal report. Then the issue still arises as to what extent does this possibly new sentence constitute the subject's actual higher order thought (cf Zimmer, 1983).

Bias Provided By Third Order Thoughts

When a person reports the content of a first order mental state, they are sometimes, but not always, conscious of that mental state. If I say the object in front of me is a circle, normally I am conscious of seeing the circle. But if I am a blindsight patient even though I say the object in front of me is a circle, I may not be conscious of seeing the circle at all. The blindsight patient reports the content of the first order mental state, but this is not accompanied by a second order thought about seeing. The same considerations apply to reporting second order thoughts. When a person reports his confidence he is expressing a second order thought. Typically, such a person will be conscious of their second order thought by virtue of having an appropriate third order thought 'I think that I am guessing', 'I think that I am seeing a circle'. When a person gives a confidence rating, they may be directly stating their second order thought, or they may be describing the second order thought as it is portrayed by their third order thought. For example, a person may be unsure what second order thoughts they are having, what state of confidence they are in ('I do not know if I am I partially seeing or if am I guessing! I think I am seeing!'); such wondering about second order states is to have third order thoughts.

Rosenthal (2000c) points out that a second order thought can be mistaken about the content of its target first order mental state. A second order thought is a representation, and as such it can misrepresent. Rosenthal gives the example of someone going to the dentist, and they expect to feel pain. As the dentist applies the drill, the person hollers in pain. The dentist tells them they couldn't have felt any pain there because he personally had removed the relevant nerve previously. The person, somehow more correctly attuning their second order thoughts to first order mental states, says 'You are right! I am just feeling vibrations!' Previously, Rosenthal argues, the person was in a first order mental state of feeling vibration, while having a second order thought about feeling pain. The person consciously felt the pain, because that was the content of the second order thought. Or consider the blindsight patient again. If he is asked whether a circle is there or not, he may simply report the content of his first order mental state and reply that there is. Or he may rely on the content of his second thought 'I see nothing' and respond that there is no circle.

Just as a second order thought can be mistaken about its target first order mental state, so a third order thought can be mistaken about its target second order thought (Rosenthal, 1995). When subjects say they are guessing, they may be correctly reporting the relevant second order thought, OR they may be reporting the representation of their second order thoughts provided by their third order thoughts. And if their third order thoughts misrepresented their actual second order thought, the subjects may be mistaken in reporting their second order thought as one of e.g. guessing. If their actual second order thought was one of seeing, then their first order mental state of seeing would have been conscious, despite subjects' reports that they were guessing. This is another way of putting the problem of reports of conscious states being biased. This bias can be a

problem for the zero correlation criterion as much as for the guessing criterion. Inaccurate third order thoughts can introduce noise in the assessment of second order thoughts; such noise will flatten the confidence accuracy relation, potentially reducing the measure of the zero correlation criterion to a level where it is unlikely to be picked up with a significance test. We need an accurate assessment of the second order thought in order to know whether the first order mental state is unconscious or not; but our measurement may be biased to some unknown degree if we end up assessing second order thoughts partly by use of fallible third order thoughts.

One might wonder how big a problem this could possibly be. First one might question whether higher order thoughts can really be mistaken about their lower order target states. The possibility of mistaken higher order thoughts strikes some people as strange. However, higher order thoughts are just representations like any other, and that means they can misrepresent. They must be products of some causal chain, some conscious or unconscious inferential process, and so there is always scope for error. Nonetheless, while there is room for error in forming a higher order thought, presumably the error is small; by and large higher order thoughts represent the lower order states quite accurately, or the higher order thoughts would not be able to perform whatever function they have that led to them coming about in the first place. Presumably, higher order thoughts normally get it right, as far as they go. But that still allows for some error.

The proportion of times a third order thought is in error about a second order thought may be less than the proportion of times a second order thought is in error about a first order state. Imagine there is some mechanism — a HOT box — that generates higher order thoughts about whatever it is given as input. Plausibly, the output of the HOT box would be available as input to it. Thus, if any first order state is accessible by the HOT box, then second order, third order, etc. thoughts also become available (Carruthers, 1996). This fits with the everyday fact that when we know something we generally know how we know it (by seeing it, hearing it, inferring it, etc.). In general, when we have a second order thought (making us conscious of some knowledge) we can readily have a third order thought (making us consciously aware of the type of mental state by which we know). Thus, while many first order states might not have access to the HOT box, any second order thought has in principle access. Thus, third order thoughts should have a high degree of accuracy relative to second order thoughts (the latter sometimes having to be formed about states not available to the HOT box). Furthermore, inaccurate third order thoughts might trigger second order thoughts consistent with the third order thoughts, overwriting the original second order thought, and bringing about greater consistency between second order and third order thoughts. Be this as it may, there is still logical scope for third order thoughts to be and to remain inaccurate, and hence scope for reports of second order thoughts to be biased.

Can we address this problem by considering how we assess a person's first order mental state independently of their second order thoughts; and then apply

the same method to assessing their second order mental states independently of their third order thoughts? On a minimal effort theory of subject performance (cf McKoon & Ratcliff, 1992), subjects will use the lowest order thought just necessary to satisfy an experimenter's request. Consistently, a good way of getting at people's first order states is to ask them to make choices amongst a range of possible first order contents. That is how we assess the first order content of blindsight patients. They say they see nothing, but we force them to choose between e.g. a number of different shapes — we don't ask about what they see (a request that requires a second order thought about seeing to answer), we simply ask them to discriminate between different possible first order contents. Similarly, in subliminal perception, if we can find any evidence of first order mental content we generally find it can be used to make forced choice discriminations (e.g. Cheesman & Merikle, 1984; Holender, 1986), even if subjects claim they can see nothing, i.e. the forced discriminations typically rely on first order content despite conflicting second order thoughts. This would be expected from the fact we are asking subjects simply about first order contents, not about their mental states. Sometimes it helps if subjects are given leave to guess and just say whatever comes to mind (Marcel, 1993), making it clear that we are not interested in the content of second order thoughts about seeing etc. There are some claims to the contrary; namely, that there can exist first order content which cannot be used in making forced choice discriminations (e.g. Greenwald *et al*., 1995) but these remain controversial (e.g. Miller, 2000). At the very least, it is difficult to find, and then replicate, conditions under which there exists first order content that cannot be used to control relevant discrimination choices (Holender, 1986; one example contrary to my claim in implicit learning is Jimenez *et al*., 1996; for the more general difficulty in finding such effects see e.g. Shanks & St John, 1994; Berry & Dienes, 1993). It is relatively easy to find conditions even where first order representations control discriminations in the presence of conflicting second order thoughts (e.g. Merikle, 1992). By the same token, a direct question for second order thoughts is answered with minimal effort by expressing any second order thought. Thus, if we give subjects a scale or range of response options representing different possible expressions of second order thoughts, and tell the subjects to say directly what comes to mind, we should be eliciting second order thoughts directly, with minimal contamination from third order thoughts. This is the procedure we follow in using the guessing and zero correlation criteria. However, we haven't thereby ruled out the possibility of measurement error in determining subjects' second order thoughts. We have not ruled out some possibility of mistaken third order thoughts affecting the measurement of second order thoughts.

Evidence For Subjective Measures Being Biased

So far we have discussed various ways in which subjective measures can be biased, creating problems for their use for measuring unconscious mental states. Is there any evidence that such bias really exists?

Kunimoto *et al*. (2001) argued that while the guessing criterion was susceptible to bias, the zero correlation criterion can be measured in a bias free way. Namely, signal detection theory can be applied to confidence ratings, seeing how high and low confidence ratings discriminate between correct and incorrect responses. A d′ measure of this discrimination can be calculated to give a bias-free measure of the zero correlation criterion.[4,5] Kunimoto *et al.* argued they provided evidence for bias in the guessing criterion, by finding the guessing criterion for unconscious knowledge could be satisfied, while the zero correlation criterion indicated the existence of some conscious knowledge.

Kunimoto *et al*. (2001 experiment one) presented subjects with one of four backward masked colour words and asked subjects to indicate after each trial which word had been presented (we will call this identity discrimination). For one group of subjects, after each block of 48 trials, subjects estimated their percent correct identity discrimination on a scale from 25% to 100%, replicating the Cheesman and Merikle (1984; 1986) procedure. Another group of subjects gave a high or low confidence judgment after each trial (and subjects had to say 'high' or 'low' about equally often). Different blocks differed according to the SOA between target stimulus and mask.

When blocks were divided into different identity discrimination accuracy bins, Kunimoto *et al.* (2001, p. 311) commented that the Cheesman and Merikle 'procedure shows no awareness for either the 0–30% or the 30–60% bin'. However, because the high-low confidence group of subjects was more accurate when the subjects said they had high rather than low confidence for the 30–60% bin, Kunimoto *et al*. argued this showed the existence of some conscious knowledge — conscious knowledge that the guessing criterion had failed to detect. Hence, they concluded, the guessing criterion is a biased measure of unconscious knowledge.

Their evidence is important, but still tentative. The statement that the 30–60% bin satisfied the guessing criterion was based on the fact that the average confidence estimate for these blocks was not significantly above 25%, just numerically so (mean = 26.4, $t(17) = 1.71$). However, confidence was above 25% in some blocks, and so the guessing criterion was not satisfied for those blocks, and hence not satisfied for the 30–60% blocks as a whole. That is, in stimulus conditions in which the zero correlation criterion provided evidence of conscious knowledge, subjects did not uniformly say they were guessing; on the other hand, the guessing criterion applies on *only* those occasions in which subjects

[4] The method of measuring the zero correlation criterion by taking the difference between average confidence when correct and average confidence when incorrect (as used by e.g. Chan, 1992; Dienes *et al.*, 1995) is not free from bias: If the subject had a liberal criterion for saying 'guess' and said 'guess' the whole time, this difference method would indicate no relationship between confidence and accuracy regardless of the real confidence accuracy relationship. On the other hand, the slope of a calibration curve (e.g. Kolb & Braun, 1995; Allwood *et al*., 2000; Dienes & Perner, 2002a; 2003) (or a difference score in percentage accuracy between high and low confidence responses) gives a measure of confidence accuracy correlation that is positive if the correlation is positive, given that the slope is defined.

[5] Note however that the zero correlation criterion is susceptible to variability in bias, as discussed above.

say they were guessing. So the guessing criterion did not strictly apply to the 30–60% bins. That is, no contradiction between the guessing criterion and the zero correlation criterion was demonstrated.

If we take all blocks (across all accuracy bins) where subjects said they were guessing, identity discrimination accuracy was above chance $t(18) = 1.77$, $p = .046$, 1-tailed. Thus, there is evidence for some unconscious knowledge in the data as a whole, according to the guessing criterion. Kunimoto *et al*. also demonstrated a positive relation between confidence and accuracy in the high-low confidence group in the data as a whole. This positive relation indicates some conscious knowledge in at least some blocks. So strictly Kunimoto *et al.* did not demonstrate that the guessing criterion is susceptible to response bias; just that both criteria together indicated the presence of both unconscious (according to the guessing criterion) and conscious (according to the zero correlation criterion) knowledge (in different blocks).

Note that this is a typical finding in artificial grammar learning.[6] Often when subjects believe they are guessing they are performing above chance; nonetheless, as their confidence increases (on other trials) so does their accuracy (e.g. Dienes *et al.*, 1995; Tunney & Shanks, in press). One can, without contradiction, take a significant guessing criterion to indicate the presence of some unconscious knowledge and a significant relation between confidence and accuracy to indicate the presence of some conscious knowledge. Typically, subjects employ both conscious and unconscious knowledge in any given experiment, so both criteria being significant can be expected as a frequent state of affairs. In fact, that is the value of having both criteria: If one only used the zero correlation criterion, the presence of a small amount of conscious knowledge could mask the simultaneous presence of unconscious knowledge.

Twyman (2001; Twyman & Dienes, submitted) attempted to manipulate the guessing and zero correlation criteria in the artificial grammar learning paradigm by manipulating variables that may change subjects' biases. Subjects gave confidence ratings trial by trial vs relative frequency judgments after blocks of trials ('What percentage of trials in that block do you think you got right?'); they were or were not trained in the meaning of probability statements; or they were told they were being underconfident or overconfident.

In the Twyman study, the zero correlation criterion and guessing criterion were quite robust. For example, the 'trial by trial confidence rating vs frequency judgments after blocks of trials' manipulation and the probability training manipulation influenced overall confidence levels but did not significantly affect the magnitude of the guessing and zero correlation criteria. On the other hand, when subjects were told that they were underconfident, their confidence increased, and the overall number of 'guess' responses decreased. For one type

[6] In the artificial grammar learning studies, confidence has typically been taken after every trial; Kunimoto *et al*., like Cheesman & Merikle, took it after every block. Block by block ratings may be less sensitive, because they allow forgetting of the mental states on particular trials, and subjects may instead determine their confidence ratings by reference to the accuracy they expect on the *sort* of trials they have been exposed to (Gigerenzer, 2000). However, see the Twyman (2001) study below.

of grammar, the percentage of correct classifications with a confidence of guessing remained unchanged. For a more salient grammar, the percentage of correct classifications with a confidence of guessing was reduced. For the more salient grammar, subjects could discern which of their 'guess' responses were more likely to be right, and give them a higher confidence, reducing the measured amount of unconscious knowledge by the guessing criterion. A change in bias to say 'guess' changed the measured amount of unconscious knowledge, conceptually replicating the claims of Kunimoto *et al*. (2001) about the sensitivity of the guessing criterion to bias.

The change in measurement of unconscious knowledge for the salient grammar in the Twyman (2001) study — from its measured presence in standard measurement conditions to its measured relative absence when subjects were warned about being underconfident — does not necessarily mean that the knowledge was always conscious. Changing instructions may have motivated subjects to search for cues they could consciously use in picking a confidence rating. For example, they might say 'That answer came to me quickly; I conclude I was not just guessing'. On Rosenthal's theory, when a second order thought is arrived at by inferences of which the person is conscious, the second order thought — even if accurate — does not render the first order state conscious. Confidence ratings arrived at by conscious inferences can lead the guessing criterion and zero correlation criterion to under-estimate the amount of unconscious knowledge (see Dienes & Perner, 2001, for discussion; Kunimoto *et al*., 2001, make a similar point). Nonetheless, it appears one of the measurements in the Twyman (2001) study (in standard conditions or after underconfidence instructions) must be in error. Or must they?

When a Change In Bias Does Not Imply Measurement Error

In interpreting the evidence for subjective measures being sensitive to bias, the locus of the bias needs to be considered. The shifting bias could be located either between the second order thought and its expression (as considered above in e.g. the effect of third order thoughts), OR it could be located between the first order state and the formation of the second order thought. Which of those locations it is makes a big difference to interpretation. From a HOT perspective, if the change in elicited confidence ratings corresponds to a change in second order thoughts, then the change corresponds to a change in the conscious status of the first order state, not simply to a change in measurement. We could call this genuine metacognitive change. If, however, the shifting bias is located between the HOT and its expression, a change in bias produces just a measurement change, and implies the potential existence of measurement error.

One form of bias is the bias postulated in signal detection theory. Bias from a signal detection theory perspective does not automatically correspond to a biased measurement of conscious status. The crucial issue is where in the chain of events the bias occurred. If the bias was an adjustable threshold used in forming the higher order thought, then an accurate measure of the higher order

thought is an unbiased measure of the conscious status of the first order state. The second order thought may be a biased or inaccurate assessment of the first order state, but it is precisely such bias that allows unconscious mental states to exist at all. If the bias occurred between forming the higher order thought and reporting what the higher order thought is, then such bias does produce inaccuracy in assessing the conscious status of mental states.

Bias can occur in other ways. Consider a first order mental state with intentional content (knowing that ..., seeing that ..., etc.). In general, such a state consists of content (portraying the world in a certain way, which Carruthers, 2000, calls *worldly subjectivity*, e.g. 'the word in front of me now is the word red') and an attitude towards that content (seeing, knowing, desiring, liking, etc.). Changing either the content or the attitude changes the mental state. Thinking that *I am in Venice* is a different mental state from wishing that *I am in Venice*; and thinking that *I am in Venice* is a different mental state from thinking *the word in front of me is the word red*.

A second order thought portrays the first order mental state in a certain way (*mental state subjectivity* in Carruthers', 2000, terms). In forming a second order thought about a first order state, error can arise in either misrepresenting the content or the attitude. If the first order state is seeing that the word in front of me is the word red,[7] then having a second order thought about e.g. liking it when the word in front of me is red does not represent the right attitude, so the seeing is unconscious. Error could also arise because the content of the first order state is portrayed in the second order thought under a different description. If the second order thought is just about seeing a pattern of dots, then the subject is not conscious of seeing the word red (even if the shape of the dots forms the word red, but this is not represented in the second order thought).

Reingold and Merikle (1993) argued that 'Statements expressing no subjective confidence may simply reflect biases introduced by either experimental instructions (i.e. demand characteristics, in psychological jargon) or an individual's preconceived ideas concerning the value of particular types of perceptual experiences for making decisions . . . a lack of subjective confidence does not necessarily indicate an absence of conscious perceptual experience' (pp. 48–9). This is a clear expression of the possible role of biases in subjective measures. But one has to still consider what type of bias is involved. There can be biases in how first order states are represented by second order thoughts (e.g. just representing that one is seeing dots), giving a person some conscious experience (that is relevant, although the person does not know it), and this is still consistent

[7] In subliminal perception the representation formed by seeing might have a minimal explicit content (explicit in the specific sense of Dienes & Perner, 1999) of just *red*, but it will implicitly convey this meaning as a feature of the individual just seen if the subject can pass objective tests about the meaning of the word. By virtue of the subject being able to discriminate above chance, there is a consumer (Millikan,1984) of the representation caused by the seeing that uses this representation to discriminate accurately. The consumer can only do this job when the representation maps onto the fact that the word in front of the subject has the meaning red. Hence, by consumer semantics, the full meaning of the representation is that the word in front of the subject has the meaning red. However, this may be represented only implicitly, as elaborated by Dienes and Perner (1999; 2002a,.b) and Perner and Dienes (1999).

with having a first order state of seeing that *the word is red* while not being consciously aware that the word is red. The seeing of the word being red is subliminal, because of, not despite, the biases.

What To Do About Biases

To summarize the argument so far, some biases are problematic for subjective measures and some are not. The critic will dwell on the fact that there is the potential for measurement bias in subjective measures (e.g. caused by third order thoughts) and this undermines their usefulness, particularly for the guessing criterion. Is this a problem? Plausibly, when we ask subjects to directly express second order thoughts, we do measure second order thoughts to a high but finite degree of accuracy. I presume that most of the time, 'guess' responses pick out second order thoughts about guessing, even though not every 'guess' response will correspond to a guess second order thought, and some second order thoughts about guessing will be expressed with a confidence of more than guessing.

This situation seems no different than almost all other measurement situations in psychology — the measurement picks out the underlying construct with a finite degree of accuracy. The important point is to use the measurements to test a theory, to show that differences in measurement correspond to predicted observable differences in the world. We may have a theory that, for example, mental states accompanied by higher order thoughts are in principle inferentially promiscuous, whereas mental states not accompanied by higher order thoughts are not inferentially promiscuous. Consistently, Merikle and Joordens (1997), in a subliminal perception paradigm, found that information satisfying the guessing criterion of consciousness was not used to plan action, although the information was highly relevant to the action; and, conversely, the information was used when it did not satisfy the guessing criterion. The (measured) second order thoughts went hand in hand with inferential promiscuity. The point is not that the mere existence of qualitative differences proved that it was the specific distinction between conscious and unconscious that held between information in the two conditions (a view criticized by Holender, 1986). The argument is the other way round: The qualitative difference is that predicted by the theory, that some information was unconscious and other information conscious, and hence the theory is corroborated (and the method has pulled its weight). Similarly, we may have a theory that the application of conscious knowledge should be more disrupted by a random number generation secondary task than the application of unconscious knowledge (based on e.g. the postulated role of a central executive or supervisory attentional system in conscious knowledge; see Perner, 2003). Consistently, in an artificial grammar learning paradigm, Dienes *et al*. (1995) found that classification decisions accompanied by a guess response were more resistant to the effects of a secondary task than decisions produced with some confidence.

A general opposing theory in both cases is that it is biases in expressing second order thoughts rather than second order thoughts themselves that result in

differences in inferential promiscuity and sensitivity to secondary tasks.[8] This opposing theory is, however, not broadly theoretically motivated, just ad hoc. Clearly, more work needs to be done in developing psychological theories of consciousness and testing the theories with subjective measures.[9] The potential for measurement bias does not preclude such a programme of research being embraced; it invites it. An interesting question becomes under what conditions do subjective measures pick different knowledge types with the properties predicted by theory. In the conditions where they work according to theory, we can regard them as measuring what we want, with measurement bias being minimal enough. In this way, conditions in which measurement bias is or is not a problem can be determined.

Conclusion

Rosenthal's (1986) higher order thought theory provides a tool by which we can see clearly the relevance of subjective rather than objective measures of the conscious or unconscious status of mental states, and also analyse the appropriate use of various subjective measures, namely, the guessing criterion and the zero correlation criterion. The higher order thought theory implies that objective measures of consciousness do not measure conscious mental states at all. Objective tests require only first order mental states for perfect performance. However, according to the theory, a mental state is rendered conscious only by second order thoughts, and there is little reason why objective discrimination accuracy of different stimuli should rely on the presence of second order thoughts. Only by attempting to measure the presence of relevant second order thoughts can the consciousness of mental states even in principle be ascertained.

Two ways of doing this are the guessing criterion and the zero correlation criterion. Both criteria can be appropriately used in determining the existence of subliminal perception and implicit learning (construed as a mechanism that produces unconscious knowledge). There is some resistance to using purely subjective measures because people's reports of second order thoughts can in principle be subject to bias of various types. A strategy for dealing with this problem is to establish how well the criteria can distinguish knowledge states in ways that would be predicted according a theory of consciousness, and therefore establish that the criteria form a useful cornerstone in a progressive research programme (Lakatos, 1970).

Because of the possible bias in elicited confidence ratings, the guessing criterion might over- or under-estimate unconscious knowledge. In the same way, the

[8] Individual experiments will of course also have individual possible counter-explanations (e.g. Kunimoto *et al*., 2001, provide one for the Merikle and Joordens, 1997, experiment).

[9] Dienes and Perner (1999; 2002a,b; 2003), Perner and Dienes (1999) and Perner (1998; 2003) provide one way of motivating relationships between conscious awareness (construed in terms of higher order thoughts), inferential promiscuity, and executive control that could be used to make predictions about the different properties of conscious and unconscious knowledge as measured by subjective measures, including the possibility of dissociations — not just associations — between higher order thoughts and executive control.

zero correlation criterion might also over- or under-estimate unconscious knowledge. For example, Dienes and Perner (2001) show that if one infers one's confidence rating on the basis of inferences of which one is conscious, this can result in the zero correlation criterion under-estimating the amount of unconscious knowledge. Conversely, Dienes and Perner (2004) show how if the perceptual system/learning mechanism creates knowledge that is mistaken according to means the system takes to be reliable, the zero correlation criterion over-estimates the amount of unconscious knowledge[10] (a source of bias that leads the zero correlation criterion to overestimate the amount of unconscious knowledge while having no such effect on the guessing criterion).

Kunimoto *et al.* (2001) argued that in practice the zero correlation criterion is likely to be less susceptible to bias than the guessing criterion. Asking subjects to discriminate between e.g. guessing and knowing means it does not matter where the criterion for saying 'guess' is placed exactly; ability to discriminate between the mental states can still be measured. As signal detection theory is the natural way psychologists think about responses being biased, having a (in signal detection theory terms) bias-free subjective measure will be appealing to psychologists. Putting aside the complications discussed above, this argument should not be taken to mean one may as well forgo the guessing criterion and just use the zero correlation criterion to detect the conscious status of mental states. Consider a case where subjects never believe they are guessing, and give high and relatively lower confidence ratings. Let us say the first order discrimination shows the subjects have knowledge, but there is no relation between confidence and accuracy. Let us also presume that this lack of relation is because when making the first order judgments, the subjects think with different degrees of certainty, but they are not aware of the precise degree of certainty with which they are thinking. Do they have unconscious first-order mental states? On the one hand, they are not precisely aware of the nature of their mental state; on the other hand, they are e.g. seeing some content and they think that they are seeing that content. On the higher order thought theory, the higher order thought, the thinking that they are seeing, would render the mental state conscious (under a description by the higher order thought that is reasonably though not completely accurate). Thus, ideally, in order to demonstrate a mental state is unconscious, one would like subjects to be unable to discriminate specifically guessing from not guessing rather than 'some degree of confidence' from 'a little more confidence'. That is, one would want the guessing criterion to be satisfied in order to demonstrate the presence of unconscious first order mental states.

It is very difficult to obtain conditions in which subjects only use unconscious knowledge. If subjects are using some mix of both conscious and unconscious knowledge, this could express itself as both the guessing and zero correlation criteria being significant. The acid test of the usefulness of the guessing criterion is then its ability to pick out knowledge that behaves according to a theory of the

[10] It turns out the problem can be overcome by showing subjects respond probabilistically with different probabilities for different items, which they do in artificial grammar learning (Reber, 1989; Dienes *et al*., 1997); see Dienes & Perner (2004) for discussion.

nature of unconscious knowledge. The presence of both conscious and unconscious knowledge in no way undermines the usefulness of either criterion. Thus, methodological concerns alone do not pick out THE criterion for measuring unconscious mental states. It is the combination of methodology with motivated theory that is needed. Each criterion has its strengths and weaknesses. To throw away all evidence for unconscious states based on a significant guessing criterion and a significant zero correlation criterion simply because the zero correlation criterion was significant may be to waste the vast majority of potential data relevant to theory testing and construction regarding the nature of conscious and unconscious states. Trying to prove simply THAT unconscious states exist (or do not exist) is as much a dead end as simply trying to prove (or disprove) the existence of ESP; in both cases, to establish the existence of the phenomenon one has to rule out the existence of either *any* conscious knowledge or *any* normal means of information transfer. In a theoretical vacuum nobody need be convinced. Progress will be made by theory development and maximizing the usefulness of data for testing theories.

The zero correlation criterion has recently been applied to theory of mind understanding in children. Ruffman *et al.* (2001) trained children on a betting task so that the bets children gave on their answers could be used as measurement of the child's confidence. Betting behaviour was very sensitive to differences in objective probabilities on a training task; nonetheless subjects bet very highly on incorrect verbal answers in the theory of mind task. Moreover, the children's eye gaze indicated knowledge of the correct answer for the theory of mind task, knowledge that was not being transmitted in the confidence ratings. Ruffman *et al.* took this to indicate the children had knowledge of theory of mind that was quite implicit. Cowey and Stoerig (1995) came close to using the guessing criterion in animals by showing that monkeys could detect stimuli in their blind field when a stimulus occurred on every trial; but when some trials contained blanks, the monkeys classified stimuli in their blind field as 'blank' rather than as 'stimuli' trials (no confidence in their accurate blind field responding?).

According to higher order thought theory, any measure of conscious awareness has face validity to the extent it plausibly measures the presence or absence of appropriate higher order thoughts. Azzopardi and Cowey (1997) argued that they demonstrated a lack of visual awareness in blindsight by showing that the d′ for making yes/no judgments ('Was a stimulus present in that time interval?') was lower than the d′ for indicating which of two time intervals the stimulus occurred in. This dissociation need not rely on the subject reporting their own state of awareness, i.e. on directly expressing a higher order thought. So does the paradigm indicate whether mental states are conscious or unconscious? In this paradigm, if the subject did have higher order thoughts about seeing whenever they did see, such thoughts should allow equivalent performance on both discrimination tasks, as the information could be used on either task, consistent with the argument of Azzopardi and Cowey. However, what would we think if subjective measures contradicted the conclusions drawn from these objective measures? Equivalent performance on the two discrimination tasks could also come

about even though the subject sincerely believed they were guessing on both tasks, i.e. the seeing was unconscious. And what if even though a dissociation was found, the subject reported vaguely seeing the stimulus and indeed, on either task, the greater the subject's confidence, the greater their accuracy (zero correlation criterion indicated conscious knowledge)? We would be forced to try to explain the results in terms of what information was used for each task instead of whether the seeing was conscious or not. This is not to undermine the importance of the Azzopardi and Cowey result; just to indicate, we have confidence that it is saying something about conscious vs unconscious seeing to the degree that the subject reported that he did not see. Indeed, Azzopardi and Cowey's result is just the sort of qualitative difference between mental states picked out as being conscious or unconscious by subjective measures that in principle allows the subjective measures to show they are pulling their scientific weight.

According to higher order thought theory, one is conscious of a first order state under a certain description; one would not be conscious of properties of the first order state that are not represented in higher order thoughts. For example, if a first order state is seeing an object moving up, but the second order thought is simply 'I know that an object moved up', then one would not be conscious of seeing. The first order state is a conscious state, but one is not conscious of it as seeing. Weiskrantz (1997) shows how subjective measures indicate this distinction applies to blindsight patients; in some cases, they see while having no conscious awareness of the stimulus in any way (no appropriate second order thoughts); in other cases, they see while consciously knowing that the stimulus had a certain property, but they have no visual experience (no higher order thoughts about seeing, just about knowing).

In all, only a few empirical studies that have been conducted on subjective measures of conscious states, and there remains much work in using them to chart the nature of the difference between conscious and unconscious mental states.

References

Allwood, C.M., Granhag, P.A., Johansson, H. (2000), 'Realism in confidence judgements of performance based on implicit learning', *European Journal of Cognitive Psychology*, **12**, pp. 165–88.

Armstrong, D. (1980), *The Nature of Mind and Other Essays* (Ithaca, NY: Cornell University Press).

Azzopardi, P. & Cowey, A. (1997), 'Is blindsight like normal near-threshold vision?', *Proceedings of the National. Academy of Sciences of the United States of America*,**25**, pp. 14190–4.

Berry, D.C. & Dienes, Z. (1993), *Implicit Learning: Theoretical and Empirical Issues* (Hove: Lawrence Erlbaum).

Block, N. (2001), 'Paradox and cross purposes in recent work on consciousness', *Cognition*, **79**, pp. 197–219.

Carruthers, P. (1992), 'Consciousness and concepts', *Proceedings of the Aristotelian Society*, Supplementary Vol. **LXVI**, pp. 42–59.

Carruthers, P. (1996), *Language, Thought, and Consciousness* (Cambridge: Cambridge University Press).

Carruthers, P. (2000), *Phenomenal Consciousness Naturally* (Cambridge: Cambridge University Press).

Carruthers, P. (2002), 'The cognitive functions of language', *Behavioral and Brain Sciences*, **25**, pp. 657–74.

Chan, C. (1992), 'Implicit cognitive processes: theoretical issues and applications in computer systems design'. Unpublished D. Phil thesis, University of Oxford.

Channon, S., Shanks, D., Johnstone, T., Vakili, K., Chin, J. & Sinclair, E. (2002), 'Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning', *Neuropsychologia*, **40**, pp. 2185–97.
Cheesman J. & Merikle, P.M. (1984), 'Priming with and without awareness', *Perception & Psychophysics*, **36,** pp. 387–95.
Cheesman J. & Merikle, P.M. (1986), 'Distinguishing conscious from unconscious perceptual processes', *Canadian Journal of Psychology*, **40**, pp. 343–67.
Cowey, A. & Stoerig, P. (1995), 'Blindsight in monkeys', *Nature*, **373**, pp. 247–9.
Dennett, D. & Kinsbourne, M. (1992), 'Time and the observer: The where and when of consciousness', *Behavioral and Brain Sciences*, **15**, pp. 183–201.
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: Oxford University Press).
Dienes, Z., Altmann, G., Kwan, L. & Goode, A. (1995), 'Unconscious knowledge of artificial grammars is applied strategically', *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, pp. 1322–38.
Dienes, Z. & Berry, D. (1997), 'Implicit learning: Below the subjective threshold', *Psychonomic Bulletin and Review*, **4**, pp. 3–23.
Dienes, Z., Kurz, A., Bernhaupt, R. & Perner, J. (1997), 'Application of implicit knowledge: Deterministic or probabilistic?', *Psychologica Belgica*, **37**, pp. 89–112.
Dienes, Z. & Perner, J. (1999), 'A theory of implicit and explicit knowledge', *Behavioral and Brain Sciences*, **22**, pp. 735–55.
Dienes, Z. & Perner, J. (2001), 'When knowledge is unconscious because of conscious knowledge and vice versa', *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society, 1-4 August, Edinburgh, Scotland. Lawrence* (Mahwah, NJ: Erlbaum Associates).
Dienes, Z. & Perner, J. (2002a), 'A theory of the implicit nature of implicit learning', in *Implicit Learning and Consciousness: An Empirical, Philosophical, and Computational Consensus in the Making?*, ed. R.M. French & A. Cleeremans (Psychology Press).
Dienes, Z. & Perner, J. (2002b), 'The metacognitive implications of the implicit-explicit distinction', in *Metacognition: Process, function, and use*., ed. P. Chambres, M. Izaute & P-J. Marescaux (Dordrecht, Netherlands: Kluwer Academic Publishers).
Dienes, Z. & Perner, J. (2003), 'Unifying consciousness with explicit knowledge', in *The Unity of Consciousness: Binding, Integration, and Dissociation*, ed. C. Frith & A. Cleeremans (Oxford University Press).
Dienes, Z. & Perner, J. (2004), 'Assumptions of a subjective measure of consciousness: Three mappings', in Gennaro (2004).
Dulany, D.E. (1997), 'Consciousness in the explicit (deliberative) and implicit (evocative)', in *Scientific Approaches To the Study of Consciousness*, ed. J.D. Cohen & J.W. Schooler (Hillsdale, NJ: Erlbaum).
Eriksen, C.W. (1960), 'Discrimination and learning without awareness: A methodological survey and evaluation', *Psychological Review*, **67**, pp. 279–300.
Gennaro, R.J. (ed. 2004), *Higher-Order Theories of Consciousness* (Amsterdam: John Benjamins Publishers).
Gigerenzer, G. (2000), *Adaptive Thinking: Rationality in the Real World* (Oxford: Oxford University Press).
Greenwald, A.G., Klinger, M.R. & Schuh, E.S. (1995), 'Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition', *Journal of Experimental Psychology: General*, **124**, pp. 22–42.
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**, pp. 1–66.
Jiménez, L., Méndez, C. & Cleeremans, A. (1996), 'Comparing direct and indirect measures of implicit learning', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, pp. 948–69.
Kelley, S.W., Burton, A.M., Kato, T. & Akamatsu, S. (2001), 'Incidental learning of real world regularities in Britain and Japan', *Psychological Science*, **12**, pp. 86–9.
Kolb, F.C. & Braun, J. (1995), 'Blindsight in normal observers', *Nature*, **377**, pp. 336–8.
Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980), 'Reasons for confidence', *Journal of Experimental Psychology: Human Learning and Memory*, **6**, pp. 107–18.
Kosko, B. (1994), *Fuzzy Thinking: The New Science of Fuzzy Logic* (Flamingo).
Kuhn, G. & Dienes, Z. (2003), 'Play it again Sam, until we like it. Can implicit learning in music go beyond the learning of adjacent chunks?', Paper presented at the meeting of the Experimental Psychology Society, Cambridge, 6–7 January 2003.
Kunimoto, C., Miller, J. & Pashler, H. (2001), 'Confidence and accuracy of near-threshold discrimination responses', *Consciousness and Cognition*, **10**, pp. 294–340.

Lakatos, I. (1970), 'Falsification and the methodology of scientific research programmes', in *Criticism and the Growth of Knowledge*, ed. I. Lakatos & A. Musgrave (Cambridge: Cambridge University Press).
Marcel, A.J. (1993), 'Slippage in the unity of consciousness', in *Experimental and Theoretical Studies of Consciousness*, Ciba Foundation Symposium 174 (Chichester: Wiley).
McKoon, G. and Ratcliff, R. (1992), 'Inferences during reading', *Psychological Review*, **99**, pp. 440–6.
Merikle, P.M. (1992), 'Perception without awareness: Critical issues', *American Psychologist*, **47**, pp. 792–5.
Merikle, P.M., & Joordens, S. (1997), 'Parallels between perception without attention and perception without awareness', *Consciousness and Cognition*, **6**, pp. 219–36.
Miller, J.O. (2000), 'Measurement error in subliminal perception experiments: Simulation analyses of two regression methods', *Journal of Experimental Psychology: Human Perception & Performance*, **26**, pp. 1461–77.
Millikan, R.G. (1984), *Language, Thought, and Other Biological Categories* (Cambridge, MA: MIT Press).
Newell, B.R. & Bright, J.E.H. (2002), 'Well past midnight: Calling time on implicit invariant learning?', *European Journal of Cognitive Psychology*, **14**, pp. 185–205.
Peirce, C.S. & Jastrow, J. (1884), 'On small differences in sensation', *Memoires of the National Academy of Sciences*, **3**, pp. 73–83.
Perner, J. (1998), 'The meta-intentional nature of executive functions and theory of mind', in *Language and Thought: Interdisciplinary Themes* , ed. P. Carruthers & J. Boucher (Cambridge: Cambridge University Press).
Perner, J. (2003), 'Dual control and the causal theory of action', in *Agency and Self-Awareness*, ed. N. Eilan & J. Roessler (Oxford: Oxford University Press).
Perner, J. and Dienes, Z. (1999), 'Higher order thinking', *Behavioral and Brain Sciences*, **22**, pp. 164–5.
Reber, A.S. (1989), 'Implicit learning and tactic knowledge', *Journal of Experimental Psychology: General*, **118**, pp. 219–35.
Redington, M., Friend, M. & Chater, N. (1996), 'Confidence judgements, performance, and practice, in artificial grammar learning', in *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, ed. G.W. Cottrell (Mawah, NJ: Erlbaum).
Reingold, E.M. & Merikle, P.M. (1993), 'Theory and measurement in the study of unconscious processes', in *Consciousness*, ed.M. Davies & G.W. Humphreys (Oxford: Blackwell).
Rosenthal, D.M. (1986), 'Two concepts of consciousness', *Philosophical Studies*, **49**, pp. 329–59.
Rosenthal, D.M. (1995), 'Multiple Drafts and Facts of the Matter', in *Conscious Experience*, ed. T. Metzinger (Exeter: Imprint Academic).
Rosenthal, D.M. (2000a), 'Consciousness, content, and metacognitive judgments', *Consciousness and Cognition*, **9**, pp. 203–14.
Rosenthal, D.M. (2000b), 'Metacognition and higher-order thoughts (extended response to commentaries)', *Consciousness and Cognition*, **9**, pp. 231–42.
Rosenthal, D.M. (2000c), 'Consciousness and metacognition', in *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference*, ed. Daniel Sperber (New York: Oxford University Press).
Rosenthal, D.M (forthcoming), 'Consciousness and higher order thought', *Encyclopedia of Cognitive Science* (Macmillan).
Rowlands, M. (2001), 'Consciousness and higher order thoughts', *Mind & Language*, **16**, pp. 290–310.
Ruffman, T., Garnham, W., Import, A. & Connolly, D. (2001), 'Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge', *Journal of Experimental Child Psychology*, **80**, pp. 201–24.
Sahraie, A., Wreiskrantz, L. & Barbur, J.L. (1998), 'Awareness and confidence ratings in motion perception without geniculo-striate projection', *Behavioural Brain Research*, **96**, pp. 71–7.
Shanks, D.R. & St John, M.F. (1994), 'Characteristics of dissociable human learning systems', *Behavioral and Brain Sciences*, **17**, pp. 367–448.
Tunney, R.J. & Altmann, G.T.M. (2001), 'Two modes of transfer in artificial grammar learning', *Journal of Experimental Psychology: Learning, Memory & Cognition*, **27**, pp. 1322–33.
Tunney, R.J. & Shanks, D.R. (in press), 'Subjective measures of awareness and implicit cognition', *Memory and Cognition*.
Twyman, M. (2001), 'Metacognitive measures of implicit knowledge'. Unpublished D.Phil thesis, University of Sussex.
Weiskrantz, L. (1986), *Blindsight: A Case Study and Implications* (Oxford: Oxford University Press).
Weiskrantz, L. (1997), *Consciousness Lost and Found* (Oxford: Oxford University Press).
Zimmer, A.C. (1983), 'Verbal vs numerical processing by subjective probabilities', in *Decision Making Under Uncertainty* , ed. R.W. Scholz (Amsterdam: North Holland).

Paper received February 2003