Subjective measures of unconscious knowledge

Zoltán Dienes

Department of Psychology, University of Sussex, Brighton, BN1 9QH.
Phone : 01273 678550
Fax : 01273 678058
Email : dienes@sussex.ac.uk

Keywords: conscious, unconscious, mental state, subjective measures, higher order thoughts, implicit learning, confidence, artificial grammar learning, serial reaction time, exclusion task

Number of Coloured Pictures : 3

Abstract

The chapter gives an overview of the use of subjective measures of unconscious knowledge. Unconscious knowledge is knowledge we have, and could very well be using, but we are not aware of. Hence appropriate methods for indicating unconscious knowledge must show that the person (a) has knowledge but (b) doesn't know that she has it. One way of determining awareness of knowing is by taking confidence ratings after making judgments. If the judgments are above baseline but the person believes they are guessing (guessing criterion) or confidence does not relate to accuracy (zero-correlation criterion) there is evidence of unconscious knowledge. The way these methods can deal with the problem of bias is discussed, as is the use of different types of confidence scales. The guessing and zero-correlation criteria show whether or not the person is aware of knowing the content of the judgment, but not whether the person is aware of what any knowledge was that enabled the judgment. Thus, a distinction is made between judgment and structural knowledge, and it is shown how the conscious status of the latter can also be assessed. Finally, the use of control over the use of knowledge as a subjective measure of judgment knowledge is illustrated. Experiments using artificial grammar learning and a serial reaction time task explore these issues.

Ever since the birth of experimental psychology at the end of the nineteenth century, psychologists have been interested in the distinction between conscious and unconscious mental states (e.g. Sidis, 1898). Recently, there has been a resurgence of interest in the distinction, as demonstrated by both purely behavioural and also brain imaging research (for example, trying to find the neural correlates of consciousness). All such research requires a methodology for determining the conscious status of a mental state. This chapter will argue for the use of 'subjective measures' for assessing the conscious status of knowledge states (see also Gaillard et al, 2006). Subjective measures measure the extent to which people think they know, as opposed to measuring how much people simply know. The assumption is that knowledge is conscious if it *subjectively* seems to people that they know when they do know, that is, if people are aware of knowing. First, we discuss the philosophical basis of subjective measures and illustrate the application of two subjective measures – the guessing criterion and the zero-correlation criterion - to learning artificial grammars. Then we will consider the problem of bias and some practical details concerning the best ways of implementing the criteria. Next we show how the criteria indicate the conscious status of only some knowledge contents (judgment knowledge) but not others (structural knowledge). We show how the conscious status of structural knowledge can also be assessed with subjective measures. Finally, we show how the control a person has over the use of knowledge can be used as an alternative subjective measure of the conscious status of judgment knowledge. In sum, the chapter should give a researcher the tools to use subjective measures in a wide variety of research settings.

Philosophical basis of subjective measures

A first-order state is a mental state that is about the world. Forming a representation in the visual system that an object is moving up is an example of a first-order representation. First-order representations allow facilitated interaction with the world, for example discriminations about the world. That is their function, that is what makes them knowledge at all. Blindsight patients, who have damage to an area of the cortex called V1, can say whether an object is moving up or down at above 80% accuracy. Yet they often claim not to be seeing, often just to be purely guessing[1] (Weiskrantz, 1997). Our strong intuition is to say the seeing is unconscious precisely because the blindsight patient is not aware of seeing; they do not have an accurate mental state about the mental state of seeing. That is, it is *because* they lack a second-order state (a mental state about a mental state) that it seems right to say their seeing is unconscious. In general, subjective measures ask people to report the mental state they are in, not just to make discriminations about the world. Subjective measures test for the presence of suitable second-order states.

One reason for urging subjective measures as the appropriate method for measuring the conscious status of mental states would be the theory that a mental state being conscious is constituted by there being (or potentially being) suitable second order states. The other reason would be the theory that a state's being conscious happens to

[1] depending on speed, at fast speeds they are aware of seeing, see also Zeki, this volume

enable higher-order states, at least for us humans. Either sort of theory justifies the use of subjective measures. We discuss both in turn.

In the 1980s David Rosenthal from New York and Peter Carruthers, then at Sheffield, independently took up an idea that can be traced back to Aristotle (in the Western tradition), namely that a mental state's being conscious arises because of actual (Rosenthal, e.g. 2005) or potential (Carruthers, e.g. 2000) higher order states. Rosenthal's argument consists of two very plausible premises. The first is that:

1. A conscious mental state is a mental state of which we are conscious.

Although this premise might seem circular, it is not. Rosenthal distinguishes two separate senses of 'conscious' used in the premise. There is *transitive consciousness*, namely consciousness *of* something. Transitive consciousness always takes an object. For example, looking at a tree and thereby being conscious of the tree. Another sense is *state consciousness* – a mental state can be conscious, as when one consciously sees, consciously thinks or consciously knows. The two senses are different. A blindsight patient is, in one way of talking, conscious of the object moving up – not consciously aware of it, but just conscious of it – because he has a visual representation that makes him sensitive to the object's motion. There is transitive consciousness (of the motion). Yet that mental state of seeing is not conscious, not state conscious. The first premise relates the two very different senses of conscious by proposing that state consciousness consists of transitive consciousness of the state. But how does transitive consciousness arise?

The second premise is that:

2. The way we become conscious of mental states is by thinking about them.

There are two ways we can become conscious of anything: By perceiving it or by thinking about it being there. I can see you there or I can close my eyes and think of you being there; either way I am conscious of you being there. Philosophers debate about whether we perceive mental states. But certainly we can think of them. (Note the theory, thus far, does not state what having a mental state – like thinking – consists in. But we do not have to solve all problems at once to be making progress.)

Putting the two premises together, Rosenthal concludes that a mental state being conscious consists of there being a higher order thought (a HOT) asserting we are in that state. (For example, we see when we are aware of seeing by thinking that we see.) If you want to deny the conclusion you need to consider which of the premises you wish to deny.

There are broadly two major philosophical intuitions concerning what consciousness consists in. One is the higher order state theory just mentioned. The other is the idea that conscious states are those that are ‘inferentially promiscuous’ or in other words ‘globally available’ (e.g. Baars, 1988). If you consciously know something you can in principle use that knowledge in conjunction with anything else you consciously know or want in order to draw inferences, make plans or form intentions. In contrast, unconscious knowledge may be available for only a limited set of uses. In adult humans it follows – given they have concepts of mental states – when you see something consciously that knowledge is quite capable of being used to make other inferences, like to think that you are seeing. The property of inferential promiscuity ensures that in adult humans conscious knowledge will enable the inference that one has that knowledge; in other words, it will enable HOTs. If when you probe for HOTs you cannot find any suitable ones, the knowledge was not inferentially promiscuous, and hence not conscious.

In summary, both major (Western) philosophical intuitions concerning the nature of consciousness – higher order state and inferential promiscuity – justify the use of subjective measures as tests of the conscious status of mental states. In higher order theories, a state being conscious is constituted by a higher order state, and in inferential promiscuity theories, a state's being conscious will allow a higher order thought if you ask for one.

Particular theories of consciousness may elaborate these themes in different ways. For example, Cleeremans (this volume) proposes a graded representation theory of consciousness. Low quality representations remain unconscious, only high quality representations may become consciousness. But importantly, Cleeremans regards representational quality as necessary but not sufficient: Conscious states also require meta-representation, i.e. one must represent oneself as having the first-order representation. Meta-representation is a higher-order state; thus, subjective measures are also directly motivated by Cleereman's theory. Lau (this volume) develops a higher-order state theory of perception and similarly urges the use of subjective rather than objective measures.

In Indian philosophy there has been a debate whether mental states are 'self-luminous' (see Gupta, 2003, pp 49-55). Self-luminosity implies 'I know' is part of each first-order cognition (a notion in Western philosophy that goes back to Descartes and that has been incorporated in some current higher-order theories, e.g. Gulick, 2004). On the other hand the Nyāya schools deny self-luminosity; cognitions are followed by higher-order cognitions making us aware of the first-order ones[2]. Whichever line of argument one takes, it follows that a conscious state allows the person to know what state they are in.

The alternative to subjective measures is objective measures. Objective measures were promoted in psychology from the 1960s onwards by those people who were skeptical about the existence of unconscious mental states. An objective measure uses the ability of a person to discriminate states of the world (for example, object moving up or down) to measure whether the mental state is conscious. When people are found able to make such worldly discriminations, the conclusion drawn is that there was conscious knowledge. But worldly discrimination only test for the existence of first-order states. It is true that a failure to make a worldly discrimination indicates the absence of conscious knowledge, but it also likely indicates absent or at least degraded unconscious knowledge (see Lau, this volume). Relying on objective measures gives a distorted picture of the nature of unconscious mental states.

The guessing and zero-correlation criteria

Knowledge is typically shown when a person makes a worldly discrimination (like ‘the object is going up’, ‘this sequence is grammatical’). To test for relevant higher order thoughts we can ask the person for their confidence in each such judgment. The

[2] However, I do not think the Nyāya school provides a higher-order state theory of conscious mental states: It is not in virtue of the higher order cognitions that the lower order ones are conscious. In fact, the higher-order cognitions simply reveal the conscious nature of the first order states (see Bhattacharya, 2003, p. 144). Nonetheless, on this approach conscious states enable higher-order states, and that is all we need to justify the use of subjective measures. I am assuming that, for the sake of argument, both schools would allow non-conscious states.

simplest confidence scale is just 'guess' if the person believes the judgment had no firm basis whatsoever, and 'know' if the person believes the judgment constituted knowledge to some extent. If on all the trials when the person says 'guess' nonetheless the discrimination performance is above baseline, then there is evidence that the person does have knowledge (performance above baseline) that she doesn't know she has (she says she is guessing). This is unconscious knowledge by the *guessing criterion*. If a person's knowledge states are conscious, she will know when she knows and when she is just guessing. In this case, there should be a relation between confidence and accuracy. Thus, a relation between confidence and accuracy indicates conscious knowledge and zero relation indicates unconscious knowledge by the *zero-correlation criterion*. Both criteria are illustrated in Figure 1 (see Dienes, 2004, and Dienes & Perner, 2001, 2004, for further discussion)[3].

***********insert Figure 1 about here***********

The criteria can be illustrated with the phenomenon of *implicit learning*, a coin termed by Arthur Reber in 1967 to indicate the process by which we acquire unconscious knowledge of the structure of an environment. An everyday example is how we learn the grammar of our native language: By age five we have learnt the main regularities, but we did not know we were learning them and could not have said what they were. (Indeed, the

[3] Note that the zero-correlation criterion should be applied by finding the relationship between confidence and accuracy within each subject (enabling one to then test the significance of the relationship over subjects). If one subject contributes just one confidence and accuracy point, the relation between confidence and accuracy may be confounded with personality variables.

mechanism is so powerful, it still beats our best attempts at conscious learning: No linguist has produced a complete grammar of any natural language.) Reber investigated implicit learning by constructing artificial grammars, i.e. arbitrary rules for determining the sequences of elements. The elements he used were letters. The strings of letters produced looked more or less random but were in fact structured. Initially, people were asked to just look at, copy down or memorise such strings for a few minutes. Then they were informed that the order of letters within each string was determined by a complex set of rules, and people classified new strings as grammatical or not. Reber found people could classify at above chance levels (typically 65% after a few minutes exposure) while being unable to freely report what the rules of the system were. Reber did not use the guessing or zero-correlation criteria as measures of the conscious status of knowledge; he used free report.

Free report is a type of subjective measure because the person normally has to believe they know something in order to report it. However, critics have been unhappy with free report as an indicator of unconscious knowledge (Berry & Dienes, 1993). Free report gives the subject the option of not stating some knowledge if they choose not to (because they are not certain enough of it); and if the free report is requested some time after the decision, the subject might momentarily forget some of the knowledge. Similarly, what the subject freely reports depends on what sort of response the subject thinks the experimenter wants. For example, if the subject classified on the basis of similarity to memorized exemplars, but thinks the experimenter wants to hear about rules, then free report may not be very informative about the subject's conscious knowledge. That is, a test must tap the knowledge that was in fact responsible for any changes in performance (the *information criterion* of Shanks and St John, 1994, and the problem of *correlated hypotheses* highlighted by Dulany, 1968). One way around the information criterion is to use confidence ratings, because then the experimenter does not need to know exactly what the knowledge is that participants use. Any knowledge the participant is conscious of using as knowledge, no matter what its content, should be reflected in the participant's confidence. Further, using confidence ratings has an advantage over free report in that low confidence is no longer a means by which relevant conscious knowledge is excluded from measurement; rather the confidence itself becomes the object of study and can be directly assessed on every trial. Indeed, Ziori and Dienes (2006) provided empirical evidence for the greater sensitivity of confidence-based methods over free report in detecting conscious knowledge. These are major benefits of the use of confidence measures of conscious knowledge.

A further strength of the zero-correlation and guessing criteria is that they do not assume that people have only conscious or only unconscious knowledge in any one condition (a desideratum of tests of conscious and unconscious knowledge repeatedly advocated by Jacoby, e.g. 1991). A guessing criterion analysis indicating the presence of some unconscious knowledge does not rule out the existence of conscious knowledge on other trials. Conversely, a zero-correlation criterion analysis indicating the presence of some conscious knowledge does not rule out the existence of unconscious knowledge in the same trials. If both criteria are statistically significant then there is evidence for both conscious and unconscious knowledge, a typical state of affairs.

In terms of the guessing criterion, Dienes, Altmann, Kwan, and Goode (1995) showed that when people believed they were guessing in the test phase of artificial grammar learning paradigm, they nonetheless classified above baseline levels. These results were replicated by Dienes and Altmann (1997), Tunney and Shanks (2003), Dienes and Perner (2003), and Dienes and Scott (2005); by Dienes and Longuet-Higgins (2003) with musical stimuli; and by Ziori and Dienes (in press) in another concept formation paradigm. In terms of the zero-correlation criterion, Chan (1992) showed subjects were no more confident in correct than incorrect decisions in artificial grammar learning. Typically, though not always, the zero-correlation criterion does indicate the presence of some conscious knowledge in artificial grammar learning. Nonetheless, Dienes et al (1995), Dienes and Altmann (1997), Allwood, Granhag and Johansson (2000), Channon et al (2002), Tunney and Altmann (2001), and Dienes and Perner (2003) replicated Chan in finding some conditions under which there was no within-subject relationship between confidence and accuracy; as did Dienes and Longuet-Higgins (2003) and Kuhn and Dienes (2006) with musical stimuli. Subjects could not discriminate between mental states providing knowledge and those just corresponding to guessing; hence, there must have existed unconscious mental states. Kelly, Burton, Kato, and Akamatsu (2001) and Newell and Bright (2002) used the same lack of relationship between confidence and accuracy to argue for the use of unconscious knowledge in other learning paradigms.

The problem of bias

In order to determine the conscious status of mental states we need to make a distinction between first order and second order states. The English language does not respect that distinction very well. If I say ‘Bill saw the tree’ I usually mean there not only was a first-order seeing of the tree but Bill was also aware of seeing the tree. Similarly, normally ‘knowing’ means awareness of knowing as well. Now let’s try to keep first-order and second-order states conceptually separate, as illustrated in Figure 2. When a person makes a judgment with a certain content, e.g. ‘this string is grammatical’, the first-order state itself may be one of guessing, one in which the system itself has no commitment to that particular content. On the other hand, the system may have a lot of commitment to the judgment (because the system has used a generally reliable method in arriving at it); the commitment may show itself in the consistency with which the same judgment is made on repeated trials, or the amount of counter-evidence or punishment needed to reverse it. In a sense, the system is ‘sure’, though it does not need to represent itself as being sure in order to be sure. That is, one can have purely first-order states of guessing, being fairly sure or certain. In English when we say the person is sure we mean also they think they are sure. But in Figure 2 the terms at the bottom are just meant to refer to first-order states and not imply the person necessarily thinks they are in those states (see Twyman & Dienes, in press).

*****insert Figure 2 about here******

The first-order states at the bottom of Figure 2 are represented in higher-order thoughts via the first mapping function illustrated. For example, perhaps when the first-order state is *guessing* and also *some confidence*, the person represents herself as just guessing. That is, there is bias in the mapping from first-order states to HOTs. This is the bias that allows unconscious mental knowledge to exist at all (Dienes, 2004); it is bias that researchers interested in unconscious states want to happen. It is precisely when the person is in a first-order state of some confidence (i.e. based on a generally reliable learning mechanism) but represents herself as guessing that we have unconscious knowledge. The manipulations that affect this mapping (motivations, rewards, types of structures to be learnt, conscious distractions, feedback on accuracy, etc) are the manipulations that affect amount of unconscious knowledge. Establishing the effect of such manipulations is one of the major tasks of research into unconscious knowledge.

The experimenter asks the subject to express their HOTs on a confidence scale, for example, a percentage scale. For a binary first-order judgment (for example, "this sequence is grammatical"/"this sequence is non-grammatical"), one could use a 50-100% confidence scale. The person gives a number between 50 and 100%. If they say 50%, it means they expect to get 50% of such answers correct, they could have well just flipped a coin. If they felt they knew to some extent, they could give a number to reflect that fact; for example 54%, meaning the person expects to get 54% of such judgments correct. And if the person was completely certain they could give 100%. Figure 2 illustrates a possible mapping from the thoughts the person actually has to the form of verbal expression allowed by the experimenter, in this case a 50-100% confidence scale (mapping function 2). Unlike the bias in the first mapping function, bias in the second mapping function is undesirable. As illustrated in Figure 3, it is possible that when people *say* they are guessing they actually *think* they have some knowledge. The knowledge demonstrated when people say they are guessing may all be due to those cases where the person thinks they do have some knowledge. Then, despite what the guessing criterion seems to indicate, there would not be any unconscious knowledge.

*****insert Figure 3 about here*****

The problem of validity is faced by all tests in psychology: Does the test measure what it says it does? Even if we solved this problem with certainty for the guessing criterion, there is also a more general point: scientists seek not to *classify* the world according to a priori criteria but to *identify* interesting kinds in nature. In the same way, the ultimate aim of the guessing criterion is not to classify knowledge according to an a priori notion of what is unconscious, but to identify an interesting kind in nature, namely, we speculate, unconscious knowledge. The evidence that it does so is provided by evidence that the criterion is useful in separating qualitatively different types of knowledge (the conscious and the unconscious) that differ specifically in ways predicted by interesting theories of the difference between conscious and unconscious. The guessing criterion's long term ability to do this is the evidence that it measures what it says it does and, more importantly, that it picks out a kind in nature that is worth studying. Notice its validation depends on its being used with substantial theories. One cannot establish whether or not there is unconscious knowledge as an isolated question in itself separate from a theoretical research program.

One common theory of conscious knowledge is that it relies on a frontal working memory system for formulating and testing hypothesis and drawing inferences, but that unconscious (implicit) learning does not. Unconscious learning, it is proposed, involves changes in the weights of neural networks that happen by automatic learning rules, requiring only minimal levels of attention to the stimuli (e.g. Berry & Dienes, 1993;

Cleeremans, this volume). If so, then loading working memory with a difficult task should interfere with the application of conscious knowledge but not unconscious knowledge. One task for loading working memory is random number generation: producing a digit (0..9), one every second or two, such that the sequence is random is very consciously demanding. Dienes et al (1995) found that random number generation during the test phase of an artificial grammar learning task interfered with the accuracy of classification when people had some confidence but not with accuracy when people believed they were guessing. There is no reason why a failure to verbalize a HOT should render knowledge resilient to the effects of a demanding secondary task; but the results are consistent with the claim that people had unconscious knowledge, and its application does not rely on working memory. On another concept formation task, Ziori and Dienes (in press) also found performance associated with guess responses was resilient to a demanding secondary task. The effects of secondary tasks on performance on implicit learning tasks more generally is variable (see Jiménez, 2003); perhaps, the effects would be clarified if subjective measures were used to separate conscious and unconscious knowledge, which has rarely been done to date (for a refinement of this claim, see the section on judgment versus structural knowledge below).

Merikle (1992) reviewed evidence that the guessing criterion picks out a qualitatively different type of knowledge in perception as well. For quickly flashed stimuli, people based plans for action on stimuli they say they saw, but not on the presence of stimuli they say they just guessed at. This provides some evidence that when people said they were guessing they also *thought* they were guessing and were not just saying it. Knowledge associated with 'guess' responses was not inferentially promiscuous, people were unwilling to base actions on it.

Skeptics of the existence of unconscious knowledge point out that unconscious rather than conscious knowledge is often associated with lower performance and so seeming qualitative differences between 'conscious' and 'unconscious' knowledge may arise from the scale effects of having different amounts of conscious knowledge (e.g. Holender, 1986). Lau (this volume) describes a perception experiment in which overall detection performance was equalized between two conditions that differed in terms of the proportion of times it seemed to the subject they saw anything. In this situation the quality of knowledge is the same for conscious and unconscious cases, an ideal method for future work exploring their qualitative differences without the confound of differing performance levels. Interestingly, in fMRI the two conditions differed in the activation of only the dorsolateral prefrontal cortex. In contrast, Spence et al (2001) found that subjects asked to lie showed increased activation in the ventrolateral prefrontal cortex, as well as many other areas (but only minimal activation of the dorsolateral prefrontal cortex). Thus the subjects in Lau's experiment were unlikely to have differed in the extent to which their words were true to their thoughts across the two conditions. That is, there does not seem to have been a bias problem in Lau's experiment.

The above arguments suggest that the guessing criterion often does track what it says it tracks: Unconscious knowledge. The bias problem, while a possible problem, is not necessarily an actual problem for the guessing criterion. That is not to say the guessing criterion will always track unconscious knowledge reliably; the conditions under which it does so is a substantial problem for future research.

The zero-correlation criterion can escape the bias problem, as illustrated in Figure 4. In this example the confidence rating is just a 'guess' or 'sure' response. In Figure 4a, the criterion separating the *verbal* response 'guess' from 'sure' occurs at the boundary between the *thought* that one is guessing and the thought that one is only more or less guessing. If the HOTs are accurate there will be a relation between confidence and accuracy. Figure 4b shows the subject has a bias that would be problematic for the guessing criterion – the criterion placement means the 'guess' verbal response includes some HOTs where the person thinks they know to some extent. BUT note: If the HOTs are accurate, there will still be a relation between confidence and accuracy. The change in bias does not affect the use of the zero-correlation criterion for indicating conscious or unconscious knowledge.

******Insert Figure 4 about here*********

One way of measuring the zero-correlation criterion is the Chan difference score (e.g. Chan, 1992; Dienes et al 1995; Ziori & Dienes, in press). For a binary confidence rating, this is the difference in the proportion of 'sure' responses when correct and when incorrect. The proportion of 'sure' responses when correct is called a *hit* in type 2 signal detection theory (STD) and the proportion of 'sure' responses when incorrect is called a *false alarm*. The Chan difference score is hence equal to hits minus false alarms, which is the commonest way of dealing with the possibility of bias in memory research.

The slope in Figure 1 is similar to the Chan difference score but it conditionalizes the other way round: P(correct/'know') minus P(correct/'guess'). This may be a better way of dealing with bias in artificial grammar learning because the slope is undefined if the subject uses only one confidence category (possibly indicating an extreme bias).

If we define *misses* as (1 – hits) and *correct rejections* as (1 – false alarms), then the following quantity:

Ln((hits X correct rejections)/(false alarms X misses))

also gives a measure of the relation between confidence and accuracy controlling for bias. If it is scaled by the factor $\sqrt{3}/\pi$, it is called (logisitic) d'("d prime"). Tunney and Shanks (2003), for example, implemented the zero-correlation criterion with d'.

In sum, the various ways of implementing the zero-correlation criterion allow the bias problem to be addressed.

The relative insensitivity of the zero-correlation criterion to bias does not imply it is better than the guessing criterion or that it should replace it. Typically, subjects presumably develop both conscious and unconscious knowledge and the use of both criteria is useful for picking these out. As mentioned earlier, the proof of the usefulness of the criteria is in their heuristic value and this has scarcely been tested yet. Also, the interpretation of the zero-correlation criterion depends on one's model of underlying processes. A lack of relation between confidence and accuracy does not automatically mean all the knowledge is unconscious. All knowledge may be unconscious if people, not being aware of any knowledge, but not wanting to give just one confidence response all the time, chose 'guess' and 'know' responses randomly. Or it may be that unconscious and conscious knowledge have the same accuracy, and the unconscious knowledge expresses itself in the 'guess' responses and the conscious knowledge expresses itself in the 'sure' responses. Indeed, if the unconscious knowledge were superior to the conscious knowledge, there may be a negative relation between confidence and accuracy. (We are currently working on a paradigm for finding just this outcome.) The criteria are not operational definitions in the literal sense of defining; they are just tools and like any tool must be used with intelligence and sensitivity in each application.

Types of confidence scales

Now we come to a practical matter. Does it matter what sort of confidence scale one uses? Tunney and Shanks (2003) compared a binary scale (high vs low) with the 50-100% scale with a particularly difficult type of artificial grammar learning task (classification performance about 55%). They found the binary scale indicated a relation between confidence and accuracy where the 50-100% scale did not, so the binary scale was more sensitive (even after a median split on the continuous scale was used to make it binary for the purposes of analysis). Tunney (2005) obtained the same result with another artificial grammar learning task (again a difficult one with performance around 55%). The result is surprising: One would think giving people more categories than two would focus their mind on finer distinctions. On the other hand, presumably HOTs are not typically expressed as numbers. The person may think something like ‘I am more or less guessing’ and the process by which this is converted into a number for the experimenter may be more variable and noisy than the process of converting it into everyday words. Still, it remains an open question whether the type of scale used does make any consistent difference to the sensitivity of the zero-correlation criterion.

In an ongoing study, I have asked different groups of subjects to express their confidence in one of six different scales: binary (high vs low); binary (guess vs sure: more useful than high vs low because it asks the subject to put the divide where we want it); numerical 50-100%, any number in the range allowed; the same again, 50-100% but with detailed explanation of what the numbers mean (i.e., as was explained above, they are expected performances); numerical categories (50, 51-59, 60-69,…., 90-99, 100); and verbal categories (complete guess, more or less guessing, somewhat sure, fairly sure, quite sure, almost certain, certain).

The first study used the same difficult materials as Tunney and Shanks (2003). The relation between confidence and accuracy was expressed in different ways: Chan difference score, d' and a correlation measure much favoured by psychologists interested in metacognition, gamma (used in e.g. Kuhn and Dienes, 2006, in an implicit learning context). The precise measure used did not affect the results at all. The results for gamma are shown in Figure 5.

******insert Figure 5 about here****

There is no indication that any scale was more sensitive than any other (perhaps surprisingly: I was betting on the verbal categories being most sensitive). Maybe the type of scale does not consistently make a major difference in this situation, but when people have more knowledge overall then surely more fine-grained scales would show their greater sensitivity. The next study, still in progress, used materials typically leading to classification performance around 65%. The results for gamma are shown in Figure 6.

*****insert Figure 6 about here*****

A one-way omnibus ANOVA comparing gamma across conditions was non-significant[4]. In sum, the evidence to date does not definitively indicate one type of confidence scale is consistently more sensitive than any other overall.

A rather different type of scale was introduced by Persaud, McLeod, and Cowey (in press). On each trial subjects chose to wager either a small amount (one pound) or a large amount of money (two pounds): If they got the trial right they received the sum wagered and if they got it wrong they lost the sum. Optimally, if one had any confidence at all one should go with the large wager. In an artificial grammar learning task, the percentage of correct decisions when a small wager was chosen was 77%, significantly above chance, indicating unconscious knowledge by the guessing criterion. The probability of a high wager after a correct decision was higher than after an incorrect decision (a Chan difference score), indicating the presence of some conscious knowledge by the zero-correlation criterion. The relation between wagering (putting your money where your mouth is) and verbal confidence ratings is an issue Persaud, Lau and myself have just started to explore.

Judgment versus structural knowledge

When a person is exposed to strings from an artificial grammar, she learns about the structure of the strings. Call this knowledge *structural knowledge*. It might consist of

[4] However, an uncorrected t-test comparing just the two conditions Tunney compared (high vs low and percentage confidence) is significant: The percentage confidence scale was more sensitive than the high/low scale.

the knowledge that an M can start a string, about whole strings that were presented, about what letters can repeat, and so on. In the test phase the structural knowledge is brought to bear on a test item to form a new piece of knowledge: the judgment, for example, that this string is grammatical. Call this knowledge *judgment knowledge*. When confidence ratings are taken, the confidence is confidence in the judgment; hence confidence ratings test for HOTs about judgment knowledge. The guessing and zero-correlation criteria test the conscious status of judgment knowledge only; that is their job. They do this job very well, but sometimes people criticize them for not testing the conscious status of structural knowledge. They say, ‘But surely unconscious knowledge might be influencing confidence ratings, so they are not a good measure of conscious knowledge’ (e.g. Allwood et al, 2000). To reword the criticism with our new concepts, it states that structural knowledge may be unconscious when judgment knowledge is conscious. This is true, but not a criticism of the guessing and zero-correlation criteria.

Consider natural language. You can tell of a sentence in your native tongue whether it is grammatical or not, be reliably right in your judgment, and be confident that you are right. You have conscious judgment knowledge. But your structural knowledge is almost entirely unconscious (try explaining to a second language learner why your version of what they were trying to say is better). When structural knowledge is unconscious and judgment knowledge is conscious the phenomenology is that of intuition. When both structural knowledge and judgment knowledge is unconscious it just feels like guessing. When people are using intuition, there may be no unconscious knowledge by the guessing criterion and a strong relation between confidence and accuracy indicating judgment knowledge is all conscious.

Dienes and Scott (2005) employed a simple way of testing for the conscious status of structural knowledge as well as judgment knowledge. In the test phase of an artificial grammar learning task, after each classification decision, as well as giving a confidence rating, subjects ticked one of four boxes to indicate the basis of their judgment: pure guess, intuition, a rule or rules they could state, or memory for part or all of a training string. The guess and intuition attributions are prima facie cases of unconscious structural knowledge and rules and memory attributions cases of conscious structural knowledge. To check subjects were identifying useful internal *kinds* by ticking boxes, two manipulations were included: First, half the subjects were informed of the rules and asked to search for them in the training phase and the other half were, as normal, not informed about rules (the first group should acquire more conscious structural knowledge than the second group); and half the subjects generated random numbers in the test phase and the other half classified with full attention (the secondary task should interfere with the application of conscious structural knowledge). Dienes and Scott found that people used the four attributions about equally often, but used the conscious structural knowledge attributions more when they had been asked to search for rules rather than just memorise; and less when they generated random numbers at test (just as one would expect). Further, the level of classification performance was above baseline for each of the attributions: There was prima facie evidence of simultaneous unconscious structural and judgment knowledge (guess attributions); unconscious structural knowledge with conscious judgment knowledge (intuition attributions; the zero correlation criterion also indicated conscious judgment knowledge in this case); and finally both conscious structural and judgment knowledge (rules and memory attributions).

Possibly measurements of the conscious status of judgment knowledge have indicated dissociations in the past because the conscious status of judgment knowledge is often confounded with the conscious status of structural knowledge. Indeed, comparing intuition with guess attributions (conscious status of structural knowledge constant, conscious status of judgment knowledge differs) revealed no differential effect of the manipulations. However, the division of knowledge into conscious and unconscious structural knowledge was relevant. Unconscious structural knowledge (classification performance based on guess and intuition attributions) was unaffected by the manipulations; but conscious structural knowledge (rules and memory attributions) was harmed by a secondary task after searching for rules. The relevant distinction for capturing a kind in nature seemed to be the difference between conscious and unconscious structural knowledge, not conscious and unconscious judgment knowledge.

Scott and Dienes (submitted) drew a similar conclusion. They found that continuous ratings of the familiarity of an item (hypothesized to reflect the continuous output of the neural network responsible for learning the grammar) predicted grammaticality classification for all structural knowledge attributions. However, when subjects searched for rules, the actual grammaticality of an item had additional predictive power above that of familiarity for only the conscious structural knowledge attributions, indicating they involved an additional source of knowledge. Again the joint in nature appeared to be between conscious and unconscious structural knowledge, not conscious and unconscious judgment knowledge.

Other recent work in my lab indicates the importance of the conscious status of structural knowledge. Riccardo Pedersini, following a study by Bierman, Destrebecqz, and Cleeremans, (2005), rewarded or punished subjects in an artificial grammar learning task after they made correct or incorrect choices. On each trial a test string was shown and skin conductance was recorded for three seconds before subjects made a response. As shown in Figure 7, and replicating Bierman et al, skin conductance was higher for incorrect than correct choices. Somehow the subjects knew when they were getting it wrong, and this created arousal, increased sweating, and hence a higher skin conductance. Pedersini also asked on each trial for a structural knowledge attribution. Interestingly, even when subjects thought they were guessing, their skin conductance revealed they knew when they correct or incorrect. A striking finding shown in Figure 7 is that for unconscious structural knowledge attributions (guess and intuition), the skin conductance separated correct from incorrect responses within the first second after a test string was shown (10 deci-seconds shown on the axes); by contrast, when structural knowledge was conscious (rules or memory), a full second was needed before correct and incorrect responses separated. In terms of this time course, the relevant distinction is between conscious and unconscious structural knowledge not conscious and unconscious judgment knowledge (guess and intuition behaved similarly). The finding is consistent with unconscious structural knowledge being embedded in the weights of a neural network so the knowledge is applied in ‘one time step’ as activation flows through the network. By contrast, application of rules and recollection often require multiple processing steps.

In sum, the interesting distinction in implicit learning paradigms may be between conscious and unconscious structural knowledge. This needs further testing and currently stands as only a hypothesis. Implicit learning research should habitually take both confidence ratings and structural knowledge attributions when first-order judgments are made. In perception, it goes without saying that structural knowledge is unconscious. The useful dividing line there – between conscious perception and subliminal perception – is between conscious and unconscious judgment knowledge.

Controlling the use of knowledge

Destrebecqz and Cleeremans (2001, 2003) used people's ability to control the use of their knowledge as a measure of its conscious status, a method developed for general use in perception and memory research by Larry Jacoby (e.g. 1991). Destrebecqz and Cleeremans used an implicit learning paradigm called the serial reaction time (SRT) task. One of four locations on a computer is indicated on each trial; the subject responds by pressing a corresponding key. From the subject's point of view it is a straightforward reaction time task. But unbeknownst to the subject the order of locations is structured; we know people learn this structure because they come to respond faster when the sequence follows the structure than when it violates it. But is the knowledge conscious or unconscious? Destrebecqz and Cleeremans asked subjects to generate a sequence. Following Jacoby's methodology, there were two conditions: Subjects either tried to generate the same sequence they had been trained on as best they could (the so-called *inclusion* condition, subjects aim to include the sequence) or to make sure they did not generate that sequence (the *exclusion* condition). When people were trying to generate the sequence (inclusion), they could do so to some extent. But both conscious and unconscious knowledge would enable this. The key finding was that when people were trying not to generate the sequence (i.e. in the exclusion condition) they nonetheless generated the sequence at above baseline levels. As consciously knowing the sequence would lead one to perform below baseline, Destrebecqz and Cleeremans concluded subjects had acquired unconscious knowledge. Further, they showed that above baseline exclusion was associated with rapid trials; when subjects could take their time, subjects excluded more effectively. With slow trials, there was a clear difference between the extent to which the sequence was generated in inclusion and exclusion. The latter results are consistent with the claim that conscious knowledge takes time to apply.

In the exclusion task, so long as subjects have keyed in to the structural knowledge, it will make them tend to generate grammatical continuations. Now they need to make a judgment before they press the key: Do they know it is grammatical or is it just a random guess? If it seems like a random guess they can go ahead and press the key. If they believe it is the product of knowledge they should withhold the response and choose another. In other words, the exclusion task is an intuitively good measure of the conscious status of knowledge because it relies on a covert assessment by the subject of whether they know. If exclusion were not controlled by HOTs, it would loose its face validity. Logically, subjects could exclude simply on the basis of pure guesses; but it would be strange to conclude from such successful exclusion that it indicated conscious knowledge when the subject denies having any knowledge whatsoever. (Indeed, Dienes et al, 1995, found that subjects by purely guessing could choose to exclude the use of one grammar and apply another in classifying strings in an artificial grammar learning experiment. Here exclusion was based on unconscious judgment knowledge.)

In excluding based on conscious knowledge, all that is required is consciously knowing whether this continuation is grammatical or not. That is, when subjects are instructed to base exclusion on conscious knowledge, exclusion only requires conscious judgment knowledge. The subject does not need to know why the continuation is grammatical. Below baseline exclusion performance is prima facie evidence of conscious judgment knowledge, but is mute about whether structural knowledge is also conscious.

Fu, Fu and Dienes (in press and in preparation) replicated the Destrebecqz and Cleeremans finding of above baseline exclusion knowledge, showing it was particularly likely early in training and with statistically noisy training sequences. Conscious judgment knowledge was shown by the difference in performance between inclusion and exclusion conditions. Fu et al showed that when people made guess attributions there was no difference between inclusion and exclusion: Both measures (control and verbal attribution) agreed in showing no conscious judgment knowledge. With intuition attributions, however, there was a difference between inclusion and exclusion, indicating unconscious structural knowledge with conscious judgment knowledge. We are currently using this paradigm to explore further the qualitative differences between conscious and unconscious knowledge.

Conclusion

For many decades research into the distinction between conscious and unconscious knowledge was regarded with suspicion. William James regarded the field of the unconscious as a ‘tumbling ground for whimsies’. As late as 1994 when I gave a talk in my department on the distinction between conscious and unconscious knowledge, afterwards a colleague told me he really liked the talk, but he wondered if I could give it without referring to consciousness. Finally things have changed and it is OK to address what must be one of the most important problems in understanding minds. Please, come and have a tumble. And I urge you to seriously consider using subjective measures – despite more than one hundred years of research it seems we have barely started in seeing how useful they might be.

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*, 165-188.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Battacharya, G. (1989). *Tarkasamgraha-Dipka on Tarkasamgraha*. Progressive Publishers, Kolkata.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove: Lawrence Erlbaum.

Bierman, D., Destrebecqz, A., & Cleeremans, **A.** (2005) Intuitive decision making in complex situations: Somatic markers in an implicit artificial grammar learning task. *Cognitive, Affective, and Behavioral Neuroscience, 5,*, 297-305

Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge University Press.

Chan (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design.* Unpublished DPhil 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*, 2185-2197.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit?. New evidence with the Process Dissociation Procedure. Psychonomic Bulletin & Review, 8: 343-350.

Destrebecqz, A., & Cleeremans, A. (2003). Temporal effects in sequence learning. In L. Jiménez (Ed.), Attention and implicit learning (pp. 181-213). Amsterdam: John Benjamins Publishing Company.

Dienes (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies*. 11, 25-45.

Dienes, Z., & Altmann, G.T.M. (1997). Transfer of implicit knowledge across domains? How implicit and how abstract? In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 107-123). Oxford: Oxford University Press.

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*, 1322-1338.

Dienes, Z., & Longuet-Higgins, H. C. (2004). Can musical transformations be implicitly learned? *Cognitive Science, 28*, 531-558.

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 Erlbaum Associates: Mahwah, NJ (pp. 255-260).

Dienes, Z., & Perner, J. (2003). Unifying consciousness with explicit knowledge. In Cleeremans, A. (Ed.) *The unity of consciousness: binding, integration, and dissociation*. (pp 214-232). Oxford University Press.

Dienes, Z., & Perner, J. (2004). Assumptions of a subjective measure of consciousness: Three mappings. In R. J. Gennaro (Ed.), *Higher order theories of consciousness*. John Benjamins Publishers: Amsterdam (pp 173-199).

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*, 338-351.

Dienes, Z., & Scott, R. (submitted). The conscious, the unconscious and familiarity.

Dulany, D. (1968). Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In T. Dixon, & D. Horton (Eds.), *Verbal Behavior and General Behavior Theory* (pp. 340-387). Englewood Cliffs, NJ: Prentice-Hall.

Fu, Q., Fu, X., & Dienes, Z. (in press). Implicit Sequence learning and conscious awareness**.** Consciousness and Cognition

Gaillard, V., Vandenberghe, M., Destrebecqz, A., & Cleeremans, A. (2006). First- and third-person approaches in implicit learning research *Consciousness and Cognition, 15*, 709-722.

Gupta, B. (2003). *Cit consciousness*. Oxford University Press.

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,* 1-23.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

Jiménez, L. (Ed.) (2003). *Attention and implicit learning*. John Benjamins: Amsterdam Kelly, Burton, Kato, and Akamatsu (2001).

Kelly, S. W., Burton, A. M., Kato, T., & Akamatsu, S. (2001). Incidental learning of real world regularities in Britain and Japan. *Psychological Science*, *12*, 86-89.

Kuhn, G., & Dienes, Z. (2006). Differences in the types of musical regularities learnt in incidental and intentional learning conditions. *Quarterly Journal of Experimental Psychology, 59*, 1725-1744.

Merikle. P. M. (1992). Perception without awareness: Critical issues. *American Psychologist, 47*, 792-795.

Newell, B. R., & Bright, J. E. H. (2002). Well past midnight: Calling time on implicit invariant learning? *European Journal of Cognitive psychology*, *14*, 185-205.

Persaud, N., McLeod, P., & Cowey, A. (in press). Post-decision wagering objectively measures awareness. *Nature Neuroscience*.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 317-327.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Shanks, D.R., & St. John, M.F. (1994). Characteristics of dissociable human learning systems. *Behavioral & Brain Sciences*, *17*, 367-447.

Sidis, B. (1898). *The Psychology of Suggestion.* Appleton and Company.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. 712 D., Zheng, Y. & Woodruff, P. W. 2001 Behavioural and functional anatomical correlates of deception in humans. *NeuroReport 12*, 2849–2853.

Tunney, R. J. (2005). Sources of confidence judgments in implicit cognition. *Psychonomic Bulletin & review, 12*, 367-373.

Tunney, R. J., & Altmann, G. T. M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 1322-1333.

Tunney, R. J., & Shanks, D.R. (2003). Does opposition logic provide evidence for conscious and unconscious processes in artificial grammar learning? *Consciousness and Cognition*, *12*, 201-218.

Twyman, M., & Dienes, Z. (in press). Are subjective measures biased? Consciousness and Cognition.

Van Gulick, R. (2004). Higher-order global states (HOGS): An alternative higher-order model of consciousness. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology*. John Banjamins.

Weiskrantz, L. (1997). *Consciousness lost and found*. Oxford University Press.

Ziori, E., & Dienes, Z. (2006). Subjective measures of unconscious knowledge of concepts. *Mind & Society, 5*, 105-122.

Ziori, E. & Dienes, Z. (in press). How does prior knowledge affect implicit and explicit concept learning? *Quarterly Journal of Experimental Psychology*,

Figure captions

Figure 1: The guessing and zero-correlation criteria.

In the example, the guessing criterion indicates the presence of some unconscious knowledge and the zero-correlation criterion indicates the presence of some conscious knowledge.

Figure 2 Relations between first-order states, HOTs and verbal report

Figure 3 The problem of bias

Figure 4 Zero-correlation criterion insensitive to criterion placement

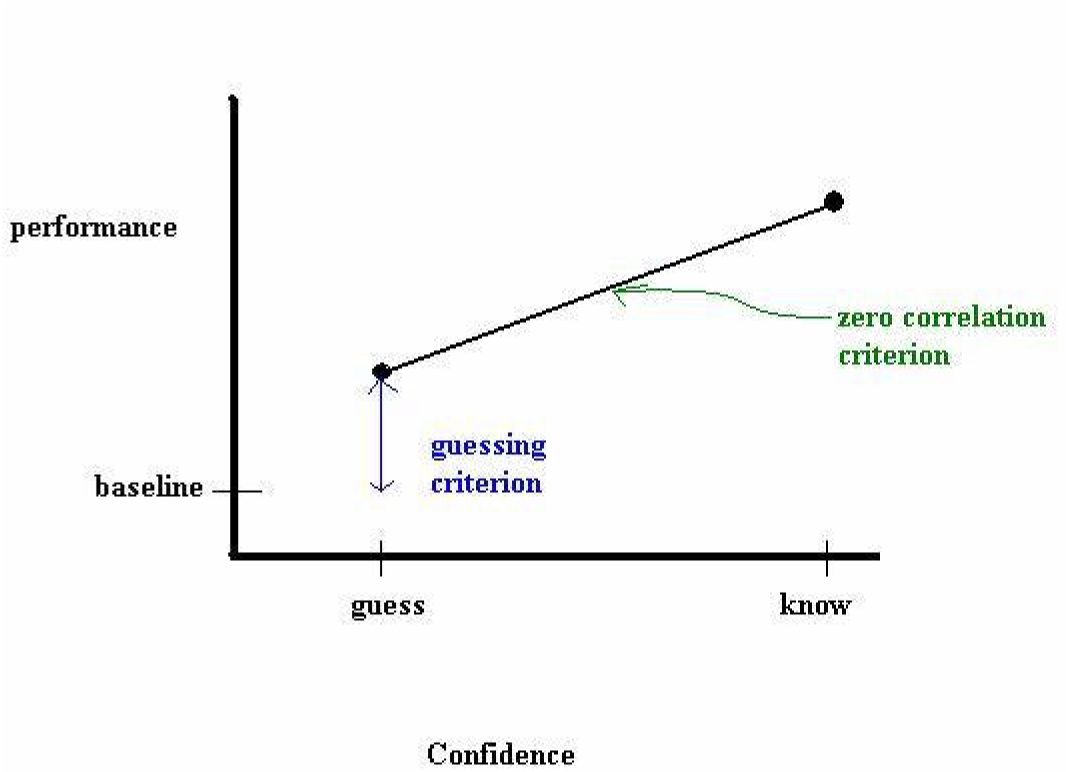
a. Good placement of criterion

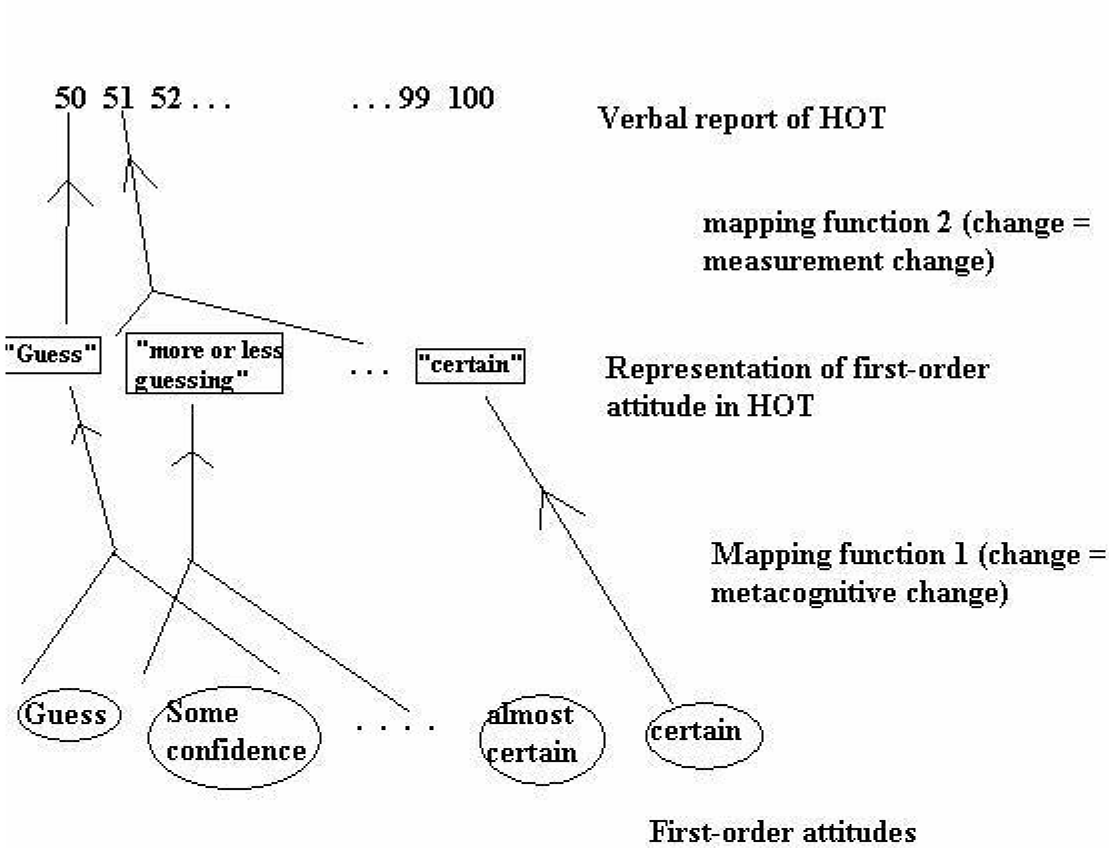
b. sloppy placement of criterion

Figure 5 Zero-correlation criterion with different confidence scales: Difficult artificial grammar learning task (overall classification performance 54%). Small squares indicate the mean for each condition, and the lines go out one standard error (SE) either side. The conditions refer to different confidence scales. For example “high versus low” is a binary scale with values ‘high confidence’ and ‘low confidence’. See text for full explanation.

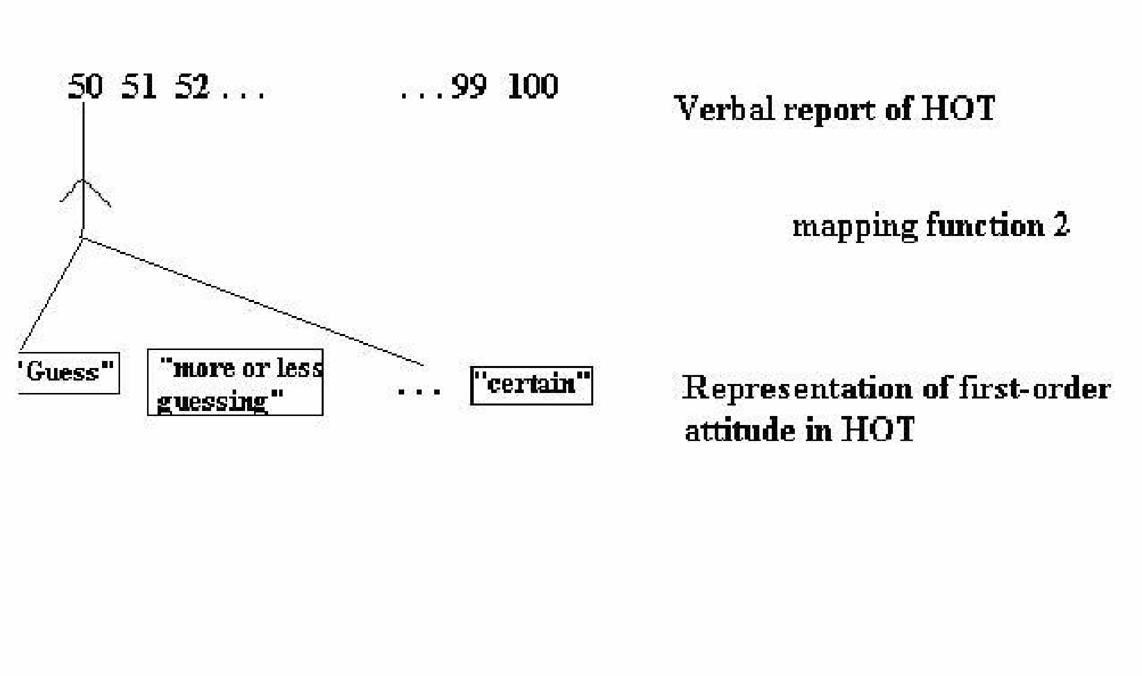
Figure 6 Zero-correlation criterion with different confidence scales: Easier artificial grammar learning task (overall classification performance 61%). Small squares indicate the mean for each condition, and the lines go out one standard error (SE) either side. The conditions refer to different confidence scales. For example "high versus low" is a binary scale with values 'high confidence' and 'low confidence'. See text for full explanation.

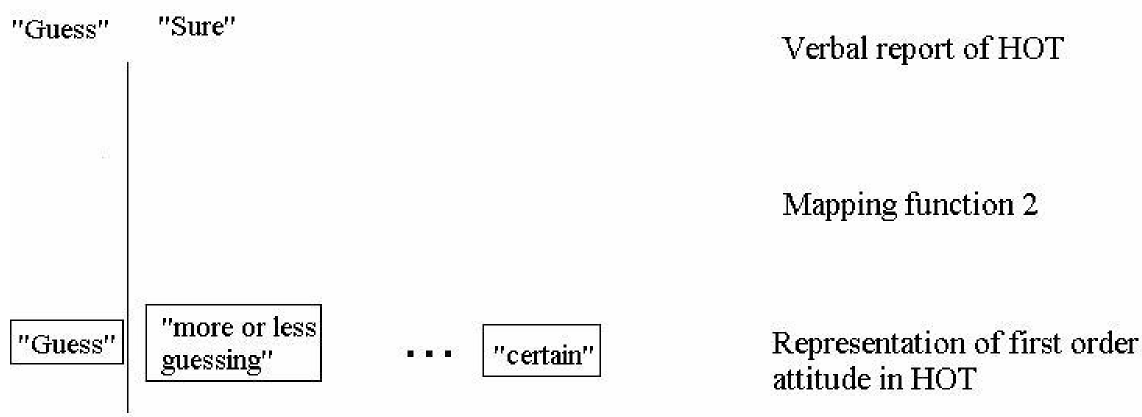
Figure 7 Galvanic Skin Response (GSR) for three seconds after a test stimulus is presented. The subject responded just as the graph finishes. The GSR measures how much the subject is sweating, i.e. arousal. The graphs show a greater GSR to incorrect than correct responses; subjects had some knowledge of when they were right, even when they thought they were guessing.

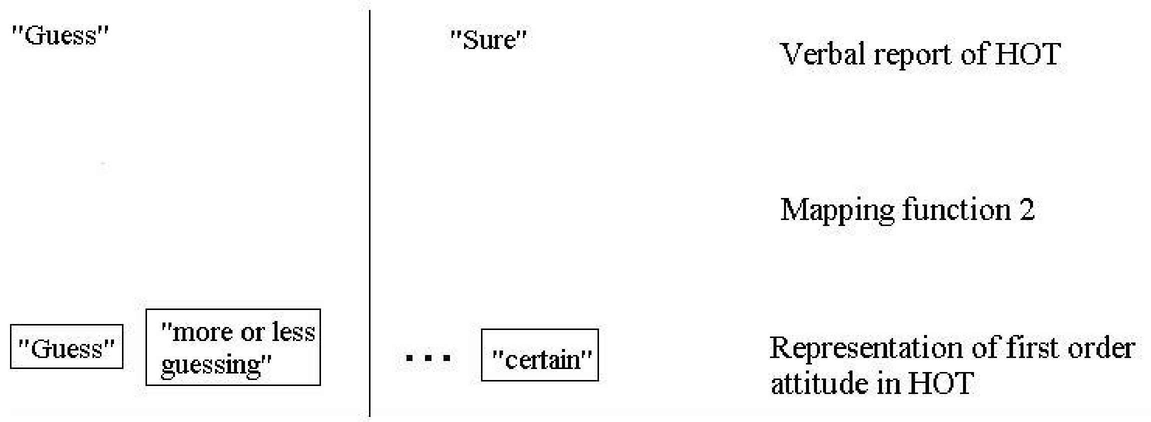

performance
zero correlation criterion
guessing criterion
baseline
guess
know
Confidence

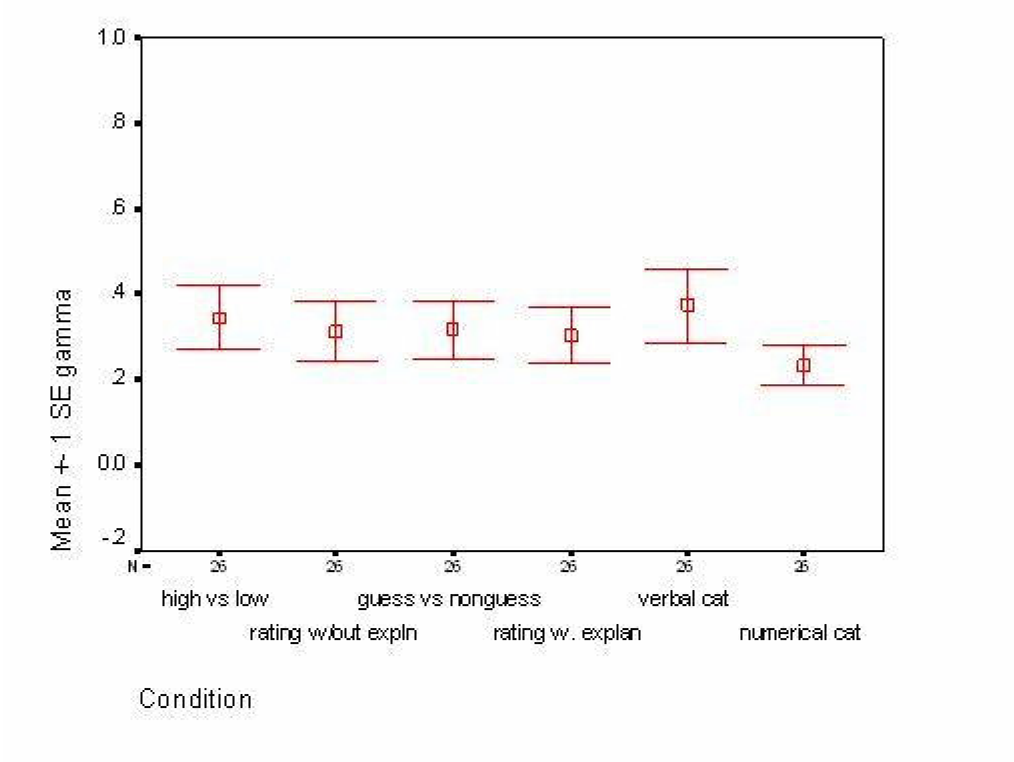


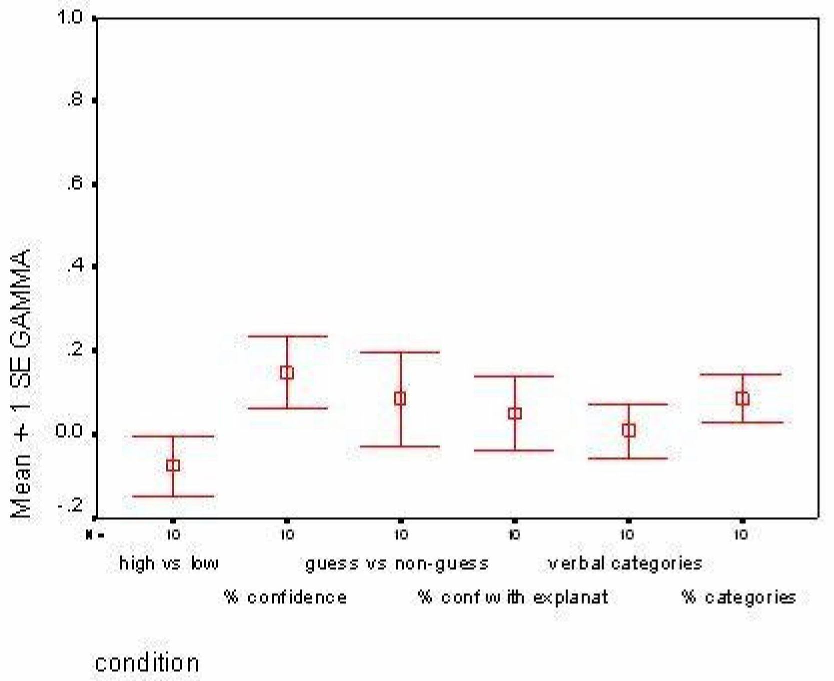
50 51 52 . . . . . . 99 100
Verbal report of HOT
mapping function 2 (change = measurement change)
"Guess"
"more or less guessing"
. . .
"certain"
Representation of first-order attitude in HOT
Mapping function 1 (change = metacognitive change)
Guess
Some confidence
. . . .
almost certain
certain
First-order attitudes

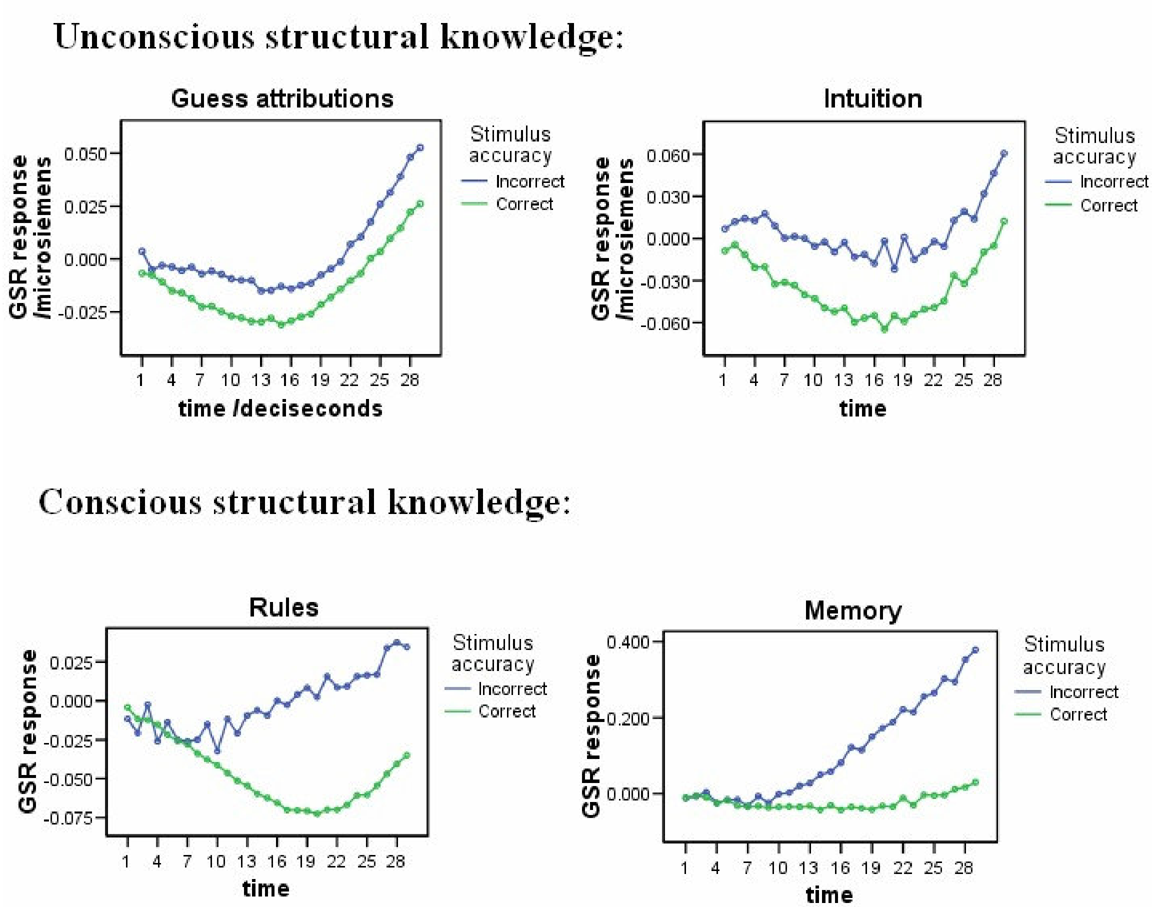

50 51 52 . . . . . . 99 100
Verbal report of HOT
mapping function 2
'Guess"
"more or less guessing"
. . .
"certain"
Representation of first-order attitude in HOT


"Guess"
"Sure"
Verbal report of HOT
Mapping function 2
"Guess"
"more or less guessing"
. . .
"certain"
Representation of first order attitude in HOT


"Guess"
"Sure"
Verbal report of HOT
Mapping function 2
"Guess"
"more or less guessing"
. . .
"certain"
Representation of first order attitude in HOT


1.0
.8
.6
.4
.2
0.0
-.2
Mean +- 1 SE gamma
N =
26
26
26
26
26
26
high vs low
rating w/out expln
guess vs nonguess
rating w. explan
verbal cat
numerical cat
Condition


1.0
.8
.6
.4
.2
0.0
-.2
Mean + 1 SE GAMMA
N =
10
10
10
10
10
10
high vs low
guess vs non-guess
verbal categories
% confidence
% conf with explanat
% categories
condition


Unconscious structural knowledge:
Guess attributions
GSR response /microsiemens
0.050
0.025
0.000
-0.025
1 4 7 10 13 16 19 22 25 28
time /deciseconds
Stimulus accuracy
Incorrect
Correct
Intuition
GSR response /microsiemens
0.060
0.030
0.000
-0.030
-0.060
1 4 7 10 13 16 19 22 25 28
time
Stimulus accuracy
Incorrect
Correct
Conscious structural knowledge:
Rules
GSR response
0.025
0.000
-0.025
-0.050
-0.075
1 4 7 10 13 16 19 22 25 28
time
Stimulus accuracy
Incorrect
Correct
Memory
GSR response
0.400
0.200
0.000
1 4 7 10 13 16 19 22 25 28
time
Stimulus accuracy
Incorrect
Correct