

The role of familiarity in implicit learning

Zoltan Dienes, Ryan B. Scott, and Lulu Wan

University of Sussex, UK

In: Higham, P., & Leboe, J. (Ed.) (2011) *Constructions of Remembering and Metacognition: Essays in honour of Bruce Whittlesea*. Palgrave Macmillan (pp 51-62).

On the conference circuit it was always an invigorating experience to be “Whittlesead”, as the saying went – to bump into Bruce at coffee, ask him a question, and hear at high speed how the issue interconnected with Bruce’s often insightful take on all the various workings of the mind. Whether or not you agreed with him, or even knew whether you agreed with him, he always provided a fresh view worthy of serious thought.

Bruce’s innovative reworking of theoretical approaches to memory included the key distinction between the production of a representation of the world, on the one hand, and an evaluation phase that determined the conscious phenomenology associated with that representation, on the other (see e.g. Mantonakis, Whittlesea, & Yoon, 2008, for a recent overview of Bruce’s ideas). We will follow this same broad distinction. That is, we accept that a mental state that just represents the world is not in itself conscious. It is only a second evaluative process that represents one as being in a certain mental state – of remembering, knowing, seeing, somehow experiencing – that allows the experience to seem a certain way, that is, to have conscious phenomenology; indeed, this assumption has been key in our own work (see e.g. Dienes, 2008, for a review).

In understanding memory and cognition, Bruce further highlighted both the relevance and irrelevance of fluency as a feature of the production phase. Fluency is the speed with which the production is formed - which when evaluated as a sign of oldness may lead to feelings of familiarity. We will also sceptically explore the role of fluency in forming people’s feelings of familiarity. In fact, the aim of this chapter is to investigate feelings of familiarity as a common feature of implicit learning, a line of inquiry that enjoys a particularly high risk of getting one Whittlesead - if only Bruce could still be found at coffee breaks on the conference circuit.

In this chapter, we will argue that familiarity refers to a uni-dimensional signal of degree of constraint satisfaction. (Bruce might regard this constraint satisfaction signal as similar to his notions of coherence or integrality - all these notions refer to global often unconscious evaluations of a production.) Familiarity is 'uni-dimensional' in that one can have more or less familiarity but it is not otherwise structured. It indicates 'constraint satisfaction' in that it indicates how predictable components of the stimulus are given past experience. The constraints so satisfied can refer to the occurrence of commonly co-occurring elements (chunks) but also to other more abstract patterns. We argue that people are often aware of this signal as a feeling of familiarity, though it can influence classification decisions unconsciously, and people are often unaware of the constraints whose satisfaction resulted in the feeling of familiarity. Despite the intuitions of many researchers, we find the signal bears no detectable relationship to fluency. Finally, and despite the intuitions of many researchers, we find familiarity can be used to control which of two bodies of implicit knowledge are used to make a classification.

I. Implicit learning and familiarity

Implicit learning is a process by which one "learns about the structure of a fairly complex stimulus environment without necessarily intending to do so and in such a way that the resulting knowledge is difficult to express" (Berry and Dienes, 1993, p 2). For example, people exposed to strings of letters generated by an artificial grammar can later recognise new strings as grammatical or not, while finding it hard to describe what distinguishes grammatical from non-grammatical strings (Reber,

1967). Artificial grammar learning is a paradigm that both we and Whittlesea have used extensively as a testing ground for theories about implicit learning.

There is some debate over what can be implicitly learned (see Pothos, 2007, for a review). Reber proposed we learn abstract knowledge; for example, the allowable bigrams and their frequency in the grammar (Reber & Lewis, 1977). Whittlesea favoured the idea that we store only the individual experiences of each training episode (e.g. Whittlesea & Dorken, 1994), continuing (and elaborating) a Canadian tradition started by Brooks (1978). Brooks proposed that people store each training item, and the similarity of a test item to training items can be used to classify without the need for explicitly represented rules, contra Reber. We have favoured neural network models of implicit learning (e.g. Dienes, 1992; Kuhn & Dienes, 2008), which can flexibly learn a number of regularities, approximate processing principles of the brain, and can have a degree of abstractness along a continuum from exemplar models (like Brooks) to symbolic representations of rules (Cleeremans, 1993).

Whatever the precise basis of the learning mechanism, modellers have found it useful to postulate a summary signal of the net coherence a test item has with the structure of the training items. For example, in neural network models of artificial grammar learning (see Cleeremans & Dienes, 2008, for a review), the network attempts to reproduce the presented test stimulus. The more the test stimulus satisfies all constraints learned from the training items, the better the reproduction. In order to classify, the correlation or match between the reproduction and the stimulus is calculated. Thus, the correlation or match is a uni-dimensional signal of degree of constraint satisfaction. The greater the signal, the greater the probability that the test item is classified as grammatical.

Brooks' (1978) key insight that people can show rule-like behaviour by learning only specific exemplars has been implemented in a number of computational models. For example, Hintzman's (1988) MINERVA model, used recently by Jamieson and Mewhort (in press) to model artificial grammar learning, stores all training items. A test item will produce an echo whose *content* reflects which parts of the test item resonate best with all training items, and whose *intensity* reflects the overall similarity of the test item to all training items. Incidentally, the echo content comprises a constant automatic construction of abstract representations (in principle unconscious until "evaluated"), in a similar way as a neural network automatically abstracts. The echo intensity is a uni-dimensional signal of degree of constraint satisfaction, also just as in the neural network case.

What does this signal represent? Across a range of different models of implicit learning, it is a qualitative non-conceptual indication of the degree to which the test item follows the same constraints as previously encountered items. It is thus a fine-grained signal of how old the test stimulus is, or how old its features collectively are, given that the constraints satisfied are those defined by old stimuli. In this sense, the signal represents events in the world: The previous meeting of the subject with the item or its features. It thus represents the familiarity of the subject with the test material.

'Familiarity' in the way just used refers to a state of affairs in the world, the fact of having been in contact with something or not. "I am familiar with the Simpsons" means I have had contact with the Simpsons. "My familiarity with Bach is not great" is a statement of the extent of my contact with Bach's music. In contrast, psychologists typically use the word 'familiarity' to refer to a mental state or process. Familiarity in this sense refers to the person's representation of the fact of them

having had prior contact with the presented material. It is *subjective* familiarity. In this sense, the constraint-satisfaction signal *is* the person's familiarity with the test item.

The signal is a *first-order* representation in the sense that the representation is about the world. As such, according to higher order theories of consciousness (Carruthers, 2007; Dienes, 2008; Rosenthal, 2002), it is an unconscious representation (or mental state) until the person is aware of being in that state - by representing to themselves that they have the first-order representation. Representing "I have a feeling of familiarity" makes the familiarity conscious. Consistent with Whittlesea's notion of an evaluation stage, it is this second-order representation that creates the *conscious* feeling of familiarity. Familiarity as a feeling, a first order representation, is the second sense of familiarity; familiarity as specifically a conscious feeling (i.e. as a second order representation) constitutes a third possible sense of the word familiarity. But we will mainly use familiarity in the second sense, subjective familiarity, the subject's take on the extent of their acquaintance with the test material, a feeling that will often be conscious but need not be¹.

Familiarity can in principle be based on many sorts of constraints: The relevant constraints cannot be determined a priori. Servan-Schreiber and Anderson (1990) were the first to suggest people may use feelings of familiarity in the artificial grammar learning task, and they suggested such feelings were based on the frequency with which chunks of letters in the test item had occurred in the training phase.

Indeed, in classifying test items, people do become sensitive to the frequency of

¹ In the second sense of familiarity, familiarity can be unconscious. This may strike some readers as odd. How can a feeling be unconscious? According to higher order theories, any mental state, including a feeling, is unconscious unless you are aware of having it. Thus, a blindsight patient can see – but the seeing is unconscious because the person is unaware of seeing. Operationally, unconscious familiarity could allow a person to discriminate which of two stimuli were old while the person believes they are literally guessing.

training chunks as suggested (e.g. Knowlton & Squire, 1994). But we now know that in learning artificial grammars people can become sensitive to other features as well: For example, the pattern of repetition structure in a string of letters (e.g. Vokey & Higham, 2005; Whittlesea & Dorken, 1994) or, more abstractly, global symmetries in musical stimuli (Kuhn & Dienes, 2005). That is, familiarity might be based on a range of possible constraints. Familiarity is a particular similarity function mapping from test items to training items, and what that function is in any particular context needs to be determined empirically. Anecdotally, it seems people have little insight into the nature of the similarity function. They don't know what or even what sort of feature might make a person or object seem familiar or unfamiliar, and have little control of its terms i.e. of the features and their weightings contributing to the feeling of familiarity. We show people are indeed often unaware of the features upon which familiarity is based. The lack of control seems so self-evident that Jacoby (1991) took uncontrollability to be definitional of familiarity. But later in this chapter we turn it into an empirical issue and show people do have some broad control over the terms of the function (even if they do not know what its terms are).

Persuasive evidence that people do rely extensively on a uni-dimensional signal in classifying test items in the artificial grammar learning paradigm comes from analyses of receiver operating characteristics (ROCs). ROCs provide an indication of the underlying cognitive processes on which judgments are based. For example, Kinder and Assmann (2000) demonstrated that the ROCs for an artificial grammar learning task are consistent with a signal detection model that assumes a continuous underlying dimension, which they postulate to be familiarity. Lotz and Kinder (2006) further demonstrated that ROCs remain consistent with a continuous underlying dimension under transfer conditions, i.e. when tested on materials employing the same

grammar but instantiated in a different letter set from the training stimuli.. Tunney and Bezzina (2007) employed ROC analyses to show that in artificial grammar learning there was not exclusive reliance on a continuous underlying dimension but such reliance increased with delay after the training phase.

If we are interested in the role of subjective familiarity in implicit learning and artificial grammar learning, we need a direct test of that role. The most direct way of measuring the extent to which a person represents their acquaintance with test material is to ask them to rate their feelings of familiarity. Higham and Vokey (2004) collected such ratings in the context of a word recognition experiment. Norman, Price, Duff, and Mentzoni (2007) did not take ratings but found that at the end of an implicit learning experiment, people often stated they used feelings of familiarity to determine their responses. In a series of papers, we asked people to rate their familiarity on a 100 point scale after each test item on an artificial grammar learning task (Scott & Dienes, 2008, 2009, in press, submitted a,b) so we could explore the determinants and the role of such feelings of familiarity in implicit learning. We found rated familiarity of test strings was predicted by various structural properties of the training strings ($R= 0.41$). Familiarity was correlated with repetition structure, the frequency of bigrams and trigrams in the training phase, and the overlap with specific training strings (in that order). That is, familiarity becomes sensitive to the structural properties of the items on which the person was trained. Further, rated familiarity strongly predicted classification responses ($r = .70$). When subjects said they were using rules or recollections, familiarity was just a partial predictor of grammaticality judgments; the actual grammatical status of the strings had additional predictive power, indicating in these cases subjects used knowledge sources other than familiarity – consistent with what subjects said they were doing. However, when subjects said, in contrast, that

they were just guessing, or using intuition or familiarity, familiarity predicted grammaticality judgments without grammatical status having any extra predictive power. In these cases, familiarity seemed to be the sole determinant of subjects' responses that allowed discrimination between grammatical and non-grammatical strings.

We will now discuss the relationships we found between familiarity and awareness, familiarity and fluency, and familiarity and control.

II Familiarity and awareness

Implicit learning occurs, in our lights, when a person has acquired unconscious knowledge - unconscious even when they are actively using it. For example, the weights in a neural network represent constraints that a person actively uses in classifying but the person may not in any way represent that they know each constraint. A person may use the constraint that 'MT' is a bigram in the training phase, but not be aware of this knowledge even as they use it to classify a string. Whittlesea has argued a different definition of implicit learning: It occurs when people learn knowledge for one purpose and do not know that they can use it for another. They may learn that MT is a bigram in the training phase, but not know that this knowledge is relevant to classifying a string as grammatical, even if they use it to do so. See Whittlesea and Dorken (1997) and Dienes and Berry (1997) for an exchange on our approaches to implicit learning. In many cases the two ways of putting it will agree. Consistent with Whittlesea's formulation, we have shown that sometimes people consciously know a rule (or consciously have a memory) but also consciously think it is not relevant and hence say that they are completely guessing

(Dienes & Scott, 2005; Scott & Dienes, 2008). However, and there may be no disagreement with Whittlesea here, people sometimes apply knowledge when they consciously believe they are producing responses literally at random: It doesn't seem to them that they are using knowledge thinking it is irrelevant, it seems to them, so they say, that they are using nothing at all (Scott & Dienes, 2008). To illustrate a stronger contrast between the two definitions, we also showed that unconscious knowledge can be formed even when it is learned for the very purpose for which it is tested: People asked to find the rules of a grammar in the training phase, can still learn to classify test items as rule-governed or not while believing they are purely guessing or using intuition, and not using memory or rules at all (Dienes & Scott, 2005; Scott & Dienes, 2008). That is, while the acquisition of unconscious knowledge can be incidental, it need not be. Further, Whittlesea's approach might imply that one can only have implicit knowledge when it is an indirect consequence of something attended to and consciously encoded as such. Yet, important as it is, attention does not completely determine the acquisition and application of implicit knowledge. People systematically misrepresent the perceptual variables controlling interception of balls, so they cannot be consciously attending to those variables as such, even as they successfully use those variables to intercept balls (Reed, McLeod, & Dienes, in press).

Subjective familiarity may be the meeting ground of believers and sceptics of unconscious knowledge. A defining feature of implicit learning is the phenomenology: When people apply their knowledge they feel like they are guessing or using intuition or familiarity. But if in all these cases people had conscious feelings of familiarity, and these conscious feelings accounted for people's ability to discriminate grammatical from non-grammatical strings, the sceptic can say "You see, there are conscious feelings, the knowledge is expressed consciously even in the

defining cases of implicit learning” (cf Berry, Shanks, & Henson, 2008; Shanks, 2005). Consistently, Reber, who has always been a strong advocate of unconscious knowledge, has long said that people know *that* they know something, they just do not know *what* they know (e.g. Reber, 1989). Norman et al (2007) expressed this sentiment by suggesting familiarity was a prime example of *fringe* consciousness: A conscious feeling acting as a pointer to the existence of unconscious knowledge. They argued such fringe feelings mediated responses on implicit learning tasks.

What does the evidence say? As indicated above, Dienes and Scott (2008) found that when people said that they were guessing or using intuition or familiarity, people’s rated feelings of familiarity accounted for grammaticality judgments. Even when people said they were responding randomly, that they were not using feelings of familiarity, their rated feelings of familiarity nonetheless accounted for their grammaticality judgments. In these cases, it seems people were not only unaware of the relevance of their feelings of familiarity, they were unaware that they were even using them. Nonetheless, people, on the face of it, had conscious feelings that contained relevant information². Both sceptics and believers can be happy.

However, if familiarity is not intrinsically conscious one might expect to find cases where unconscious familiarity controls performance. We present evidence for two such cases. Scott and Dienes (submitted) used a transfer paradigm where people were tested on a different letter set than they were trained on (but using the same grammar). In this case, we found that it was only when people said they were

² Being conscious of using familiarity does add something to using it only unconsciously: In Experiment 3 of Scott and Dienes (2008), when people said they were using familiarity, the average rated familiarity on a 100-point scale was 56 (SE = 2.5), significantly higher than the average familiarity when people said they were responding literally randomly (47, SE = 3.6) or even than when people said they were using intuition (46, SE = 2.5), p ’s < .05. More importantly, the mean correlation between familiarity ratings and grammaticality judgments was higher when people said they were using familiarity ($r = 0.56$, SE = .07) than when they said they were responding randomly ($r = .33$, SE = .10), which did not differ from when people said they were using intuition ($r = .38$, SE = .07). Thus, consciously using familiarity involves more use of feelings of familiarity than its unconscious use.

responding randomly that they performed above chance (60%); when people said they were using any other basis such as rules or recollection they were responding at chance. Importantly, when people said they were responding randomly, their rated feelings of familiarity did not completely account for their ability to discriminate grammatical from non-grammatical items; the grammatical status of the items had predictive power above and beyond rated familiarity in accounting for grammaticality judgments. What other source of knowledge could people have been using? People claim not to be using any source at all. We suggest (because we believe that implicit learning is based on a neural network) that the source is unconscious familiarity, familiarity not expressed in people's ratings. Indeed, we showed that beyond rated familiarity it was the presence of novel bigrams and trigrams that predicted people's grammaticality judgments and accounted for their additional accuracy above that achievable through rated familiarity. Novel chunks would be expected to influence familiarity; but they didn't in this case influence conscious familiarity. Why does unconscious familiarity outperform conscious familiarity in this case? We will argue below that the similarity function defining familiarity can be changed by the subject's focus of attention. It may be that when subjects allow their responses to happen "at random" the similarity function is less influenced by conscious hypotheses that, given the difficulty of the transfer task, are irrelevant or wrong.

Wan, Dienes, and Fu (2008) presented evidence for the control of the similarity function defining familiarity (we discuss this study further below). When people were trained on two grammars, they could successfully choose to endorse one or other grammar. The important point for here is that while rated familiarity distinguished the grammars when people said they were using familiarity to do so, it did not when people said they were literally guessing. Again in this case, we postulate

people were guided by unconscious familiarity, the uni-dimensional output of neural networks primed to be differentially biased to one grammar rather than another. Just as allowed by higher order theories of consciousness (a version of which Whittlesea subscribes, as noted above), people can be guided by first-order representations alone.

Whether or not the feeling of familiarity is conscious, the knowledge on which familiarity is based can be conscious or unconscious. Whether or not this knowledge consists of exemplars, processing episodes or something more abstract, we have asked subjects to indicate trial by trial whether they are aware of the basis of their judgments. We have consistently found that people use familiarity even when claiming they have no idea of the basis of their grammaticality judgments, nor of the basis of their familiarity judgments. The description of familiarity as a fringe feeling by Norman et al (2007) seems apt. It can point to unconscious structural knowledge. Nonetheless Scott and Dienes (2009) suggest that, as exposure to training and test material continues, people can use differential familiarity of parts of a string as a basis for formulating hypotheses with a separate system capable of dealing with hypotheticals. In this way, familiarity can be used to gradually explicate one's structural knowledge (cf Matthews et al, 1989). (On multiple systems, Scott & Dienes, 2009, could be contrasted with e.g. Whittlesea & Dorken, 1994, 1997.)

III Familiarity and fluency

A dominant view is that familiarity is based on fluency of processing, i.e. the speed with which it takes processing to complete. Jacoby and Dallas (1981) suggested that when processing an item with relative ease, people may attribute this to having

encountered the item before and experience it as familiar. Whittlesea and Williams (2000) elaborated this position by arguing that it was not fluency itself but surprising or discrepant fluency that can lead to feelings of familiarity.

Applying the fluency account of familiarity to models of implicit learning requires that an additional assumption be made. For example, in the context of connectionist models it requires the assumption that the constraint satisfaction signal be correlated with the speed that the network arrives at its judgment. While there is no compelling a priori reason why this assumption should be true (Bullinaria, 1994), it has proved effective in modelling human responding in the sequential reaction time task (Cleeremans, 1993). However, the implicit knowledge expressing itself in judgment tasks (like artificial grammar learning) may be quite different from implicit knowledge expressing itself in perceptual motor tasks like the sequential reaction time task (Seeger, 1994). Further, fluency is just one a priori possible basis for familiarity. Fluency is a property of the representational vehicle of one's knowledge: How fast the representation takes to form. Another vehicle property that could be used as a basis for familiarity is the total amount of activation. Conversely, familiarity may be carried purely by the content of a representation, rather than any content-irrelevant vehicle property. Indeed, in most computational models of implicit learning, processing fluency is literally the same for grammatical and non-grammatical stimuli, if not longer for grammatical rather than non-grammatical items in chunking models (Boucher & Dienes, 2003). In typical models (see Cleeremans & Dienes, 2008), familiarity is carried purely by a representation having the required content. Nonetheless, there has been a persistent assumption in the artificial grammar literature that item familiarity equates to fluency, more specifically perceptual fluency (e.g. Buchner, 1994).

Buchner (1994) employed a perceptual clarification task to explore the potential relationship between perceptual fluency and responding in AGL. The perceptual fluency of test strings was assessed based on the speed with which participants were able to make out the strings as they gradually emerged from behind a mask. The results provided mixed evidence; while grammatical strings were identified more rapidly than ungrammatical strings these differences in fluency were not found to correlate with participants' grammaticality decisions. Further the grammatical and non-grammatical strings were not counterbalanced and relevant item properties that would plausibly influence fluency were not controlled.

A relationship between perceptual fluency and judgments of grammaticality has been observed under conditions where fluency has been artificially manipulated so causal conclusions can be drawn. Kinder et al. (2003) employed a perceptual clarification task similar to that used by Buchner but manipulated perceived fluency by having half the strings clarify at a faster rate. Consistent with an influence of fluency, faster clarifying strings were endorsed as grammatical more often than slow clarifying strings and the authors concluded that fluency was the default mechanism for making grammaticality judgments.

Scott and Dienes (in press) showed that with counterbalanced grammatical and non-grammatical items, naturally occurring fluency as measured in Buchner's perceptual clarification task was not related to grammaticality. Further, fluency bore a tiny albeit significant relation to rated familiarity ($r = .08$). Familiarity was related to various structural variables, but in a way that was independent of measured fluency. In addition, when fluency was manipulated, following the same procedure as Kinder et al, it only influenced familiarity and grammaticality decisions when subjects were given limited time to make judgments (as in Kinder's procedure). Otherwise, in the

normal conditions of artificial grammar learning experiments, even fluency manipulated to be 15 times the difference observed to naturally occur between Buchner's grammatical and non-grammatical stimuli, influenced neither familiarity ratings nor grammaticality judgments. Scott and Dienes (submitted a) further showed that the limited influence of fluency extends to transfer conditions; fluency again bears a tiny, albeit significant, relation to rated familiarity ($r = .04$) and is unrelated to grammaticality. In sum, fluency carries no useful information in the artificial grammar learning task and people barely use it either for grammaticality judgments or as a basis of feelings of familiarity. These conclusions are consistent with Whittlesea and Leboe (2000) who showed with a different classification task and fluency manipulation that when participants can exploit either fluency or structural similarity to make classification judgments that they reliably favour the latter (cf also Johansson, 2009). Perceptual fluency is a *dumb heuristic* influencing responding only in the absence of actual implicit knowledge (Higham, unpublished).

IV Familiarity and control

In an elegant series of experiments, Whittlesea and his colleagues have shown that the structural level to which a subject attends (e.g. letters, syllables, words) changes the structural properties to which a person is sensitive in classifying stimuli (Whittlesea & Dorken, 1993; Whittlesea & Wright, 1997). Higham and Vokey (2004) showed that the duration with which an item is briefly displayed as a prime before presenting it for judgment influences rated familiarity. That is, it is likely that the constraints determining familiarity can be influenced by the focus of attention at

training and test and by other manipulations independent of the training phase. These considerations lead one to predict that familiarity should be partially controllable, even if one is unaware of its basis: The structural features upon which it is based may shift with context. Indeed, connectionist networks are notoriously context-sensitive, so the constraint satisfaction signal should be sensitive to context. Thus, imagining one context may make the items from that context differentially familiar.

However, the dominant approach to familiarity does not allow the conclusion that familiarity could be controllably higher for one context rather than another. Jacoby (1991) defines familiarity to be that memorial process that does not allow control. In that light, the artificial grammar learning literature contains an intriguing set of results: Dienes et al (1995) found that when subjects are trained on one artificial grammar after another, they can choose almost perfectly which grammar to use in endorsing grammatical items. That is, they can choose which context – the first or second set of training strings – is used in determining their choices in the test phase. On the Jacoby approach, familiarity, by definition, can not be used to achieve this discrimination between the grammars. Scott and Dienes (2008) by contrast found that people use, and say they use, familiarity as the main means for distinguishing grammatical from non-grammatical items when trained on one grammar. So can familiarity be used to discriminate two grammars or can't it?

Wan et al (2008) trained people on two grammars and took familiarity ratings, plus attributions from subjects concerning the basis of their judgments (random, intuition, familiarity, recollection, rules). They found that people could discriminate between the grammars almost perfectly and they could do so whatever the perceived basis of their judgment. When people say they are using recollections or rules, it is not surprising that they can discriminate. However when people say they are using

familiarity they can also discriminate. Crucially, when they say they are using familiarity they almost exclusively pick the string with the highest rated familiarity. Further, the rated familiarity for test strings consistent with their chosen grammar was greater than that for strings from the other (counter-balanced) grammar. Familiarity, subjectively defined, is sensitive to intentions and can play a key role in strategic control.

We found no tendency for the to-be-ignored grammar to be endorsed more than entirely non-grammatical items: No evidence for familiarity in the Jacoby sense. In contrast, Higham, Vokey, and Pritchard (2000) using more confusable artificial grammars found substantial endorsements of a to-be-ignored grammar compared to non-grammatical strings: On the Jacoby approach these confusions are based on familiarity. This result allows an interesting contrast between familiarity subjectively defined as we do and the Jacoby approach. The Jacoby approach would classify knowledge as based on a familiarity process even if subjects stated on each trial that their response was based on conscious rules or recollection, and discrimination between the grammars failed simply because the rules were not perfectly discriminating. Thus, a mental state is one of familiarity or not depending on a specific state of affairs external to the subject, namely, on whether some strings are classified as grammar A or B according to an experimenter's grammars - which might have been different and thus could have been classified otherwise. This is an example of externalism in the philosophy of mind (see e.g. Wilken, Bayne, & Cleeremans, 2008, entries on 'externalism' and 'representations, problems'). If the experimenter happened to classify all the new strings in the same way as the subject, there would be no familiarity. According to our approach, a state being one of subjective familiarity depends not on how the experimenter classifies strings, only on

the mechanisms operating within the subjects' mind (namely, only on whether a mechanism was used that represents oldness, regardless of whether this mechanism makes the same choices as the experimenter). While this argument biases our own preferences for a definition of familiarity, it should not be obvious that it should. Externalism as a view concerning the content of mental states is the current dominant (though not exclusive) view amongst philosophers. Jacoby's definition applies externalism in an interestingly different way, to the type of mental state one is in independent of its content; in fact, by Jacoby's approach, the very conscious status of one's mental states can be so externally determined.

V Conclusion

In summary, we hope we have drawn out arguments for the role of familiarity in implicit learning that interweave with Whittlesea's ideas in interesting ways (even if in ways he wouldn't approve). But hopefully that makes our conclusions surprisingly easy on the mind given what they are: That familiarity is typically conscious but can be unconscious; that it is often a conscious fringe feeling signalling that there exists further unconscious knowledge; that it might bring sceptic and believer together in harmony; that in cases where one has structural knowledge fluency plays almost no role in determining familiarity; and that familiarity can be controlled by intentions and used to control choice of items from different contexts, even as one remains unaware of its basis.

References

- Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove: Lawrence Erlbaum.
- Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008). A single-system account of the relationship between priming, recognition, and fluency. *Journal of Experimental Psychology-Learning Memory and Cognition*, 34(1), 97-111.
- Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and concepts*. Hillsdale, NJ: Erlbaum.
- Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, 120(3), 316-323.
- Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, 27(6), 807-842.
- Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 550-566.
- Bullinaria, J. A. (1994). Modelling Reaction Times. In L. S. Smith & P. J. B. Hancock (Eds.), *Neural Computation and Psychology - Proceedings of the 3rd Neural*

Computation and Psychology Workshop, Stirling, 31st August 1994 (pp. 34-48). New York: Springer.

Carruthers, P. (2007). Higher order theories of consciousness. In Velmans, M., & Schneider, S. (Eds) (2007). *The Blackwell Companion to Consciousness*. Blackwell (277 – 286).

Cleeremans, A. (1993). *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing*. Cambridge, MA: MIT Press.

Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning. In R. Sun (Ed.), *Handbook of Computational Cognitive Modeling* (pp. 396 - 421). Cambridge: University Press.

Dienes, Z. (2008) Subjective measures of unconscious knowledge. *Progress in Brain Research, 168*, 49 - 64.

Dienes, Z., & Berry, D. (1997). Implicit synthesis. *Psychonomic Bulletin and Review, 4*, 68-72.

Dienes, Z., & Scott, R. B. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*(5-6), 338-351.

Higham, P. A. (unpublished manuscript). Dumb Use of the Fluency Heuristic: Reliance on Perceptual Fluency in the Absence of Perceptual Facilitation.

Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *57A*(4), 714-744.

Higham, P. A., Vokey, J. R., & Pritchard, J. (2000). Beyond dissociation logic: Evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General*, *129*(4), 457–470.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306-340.

Jamieson, R. K., & Mewhort, D. J. K. (in press). Applying an exemplar model to the artificial grammar task: String-completion and performance on individual items. *Quarterly Journal of Experimental Psychology*.

Johansson, T. (2009). In the fast lane toward structure in implicit learning: Non-analytic processing and fluency in artificial grammar learning. *European Journal of Cognitive Psychology*, 21(1), 129-160.

Kinder, A., & Assmann, A. (2000). Learning Artificial Grammars: No evidence for the acquisition of rules. *Memory and Cognition*, 28(8), 1321-1332.

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, Fluency, and the Explicit/Implicit Distinction in Artificial Grammar Learning. *Journal of Experimental Psychology: General*, 132(4), 551-565.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 79-91.

Kuhn, G., & Dienes, Z. (2005). Implicit Learning of Nonlocal Musical Rules: Implicitly Learning More Than Chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1417-1432.

Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition*, 106, 184-206.

Lotz, A., & Kinder, A. (2006). Transfer in Artificial Grammar Learning: The Role of Repetition Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 707-715.

Mantonakis, A., Whittlesea, B. W. A., & Yoon, C. (2008). Consumer Memory, Fluency, and Familiarity. In C. P. Haugtvedt, P. Herr, & F. R. Kardes (Eds) *The Handbook of Consumer Psychology*. Lawrence Erlbaum Associates. (pp. 77-102).

Mathews, R. C., Buss, R. R., Stanley, W. B., & Blanchard Fields, F. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1083-1100.

Norman, E., Price, M. C., Duff, S. C., & Mentzoni, R. A. (2007). Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition*, 16, 809-837.

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855-863.

Reber, A. S., & Lewis, S. (1977). Implicit learning: an analysis of the form and structure of a body of tacit knowledge. *Cognition*, 114, 14-24.

Reed, N., McLeod, P., & Dienes, Z. (in press). Implicit knowledge and motor skill: What people who know how to catch don't know. *Consciousness & Cognition*.

Rosenthal, D. M. (2002). Consciousness and higher order thought. *Encyclopedia of Cognitive Science*. Macmillan Publishers Ltd (pp. 717-726).

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology-Learning Memory and Cognition*, 34(5), 1264-1288.

Scott, R. B., & Dienes, Z. (2009). The metacognitive role of familiarity in artificial grammar learning: transitions from unconscious to conscious knowledge. To appear. In A. Efklides & P. Misailidi (Eds.), *Trends and Prospects in Metacognition Research*: Springer, US.

Scott, R. B., & Dienes, Z. (in press a). Fluency does not express implicit knowledge of artificial grammars. *Cognition*.

Scott, R. B., & Dienes, Z. (submitted a). Knowledge applied to new domains: the unconscious succeeds where the unconscious fails. *Consciousness and Cognition*.

Scott, R. B., & Dienes, Z. (submitted b). The influence of prior familiarity on implicit knowledge and the conscious status of structural knowledge. *Consciousness and Cognition*.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115(2), 163-196.

Servan Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 592-608.

Shanks, D. R. (2005). Implicit Learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (pp. 202-220). London: Sage.

Tunney, R. J., & Bezzina, G. (2007). Effects of retention intervals on receiver operating characteristics in artificial grammar learning. *Acta Psychologica*, *125*(1), 37-50.

Vokey, J. R., & Higham, P. A. (2005). Abstract Analogies and Positive Transfer in Artificial Grammar Learning. *Canadian Journal of Experimental Psychology*, *59*(1), 54-61.

Wan, L. L., Dienes, Z., & Fu, X. L. (2008). Intentional control based on familiarity in artificial grammar learning. *Consciousness and Cognition*, *17*(4), 1209-1218.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*(2), 227-248.

Whittlesea, B. W. A., & Dorken, M. D. (1997). Implicit learning: Indirect, not unconscious. *Psychonomic Bulletin and Review*, *4*(1), 63-67.

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General*, *129*(1), 84-106.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 547-565.

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and Explicit) Learning: Acting Adaptively Without Knowing the Consequences. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23(1), 181-200.

Wilken, P., Bayne, T., & Cleeremans, A. (Eds.). (2008). *The Oxford companion to consciousness*. Oxford University Press.