Conscious And Unconscious Mental States

Zoltan Dienes and Anil K. Seth

In: O. Braddick (Ed.), *Oxford Research Encyclopedia of Psychology*, Oxford University Press. <https://oxfordre.com/psychology>

Summary

The major theories of consciousness that distinguish conscious from unconscious states can be grouped into two main classes, either higher order or integration theories. There is evidence that different types of mental states can be unconscious, though that conclusion depends on the theory of consciousness assumed. Unconscious memory (in the sense of the influence of a prior event not recollected) can shape perception, liking, and control our behaviour. Subliminal perception can produce semantic priming, guide attention and decision making; and optical variables a person describes incorrectly can guide action. Implicit learning can shape judgments and choices in complex environments. Unconscious intentions can allow people to respond appropriately in goal directed ways, while the person experiences the actions as involuntary. Unconscious attitudes are no more or less plausible than any other mental state being unconscious, but it has been hard to obtain evidence for unconscious attitudes, as distinct from gut reactions one does not agree with.

Keywords:

unconscious mental state, implicit memory, subliminal perception, implicit learning, hypnosis, implicit attitudes, subjective threshold, higher order theory, global workspace theory

Theories of Consciousness

Theoretical approaches to explaining consciousness can be grouped into ‘higher-order’ theories (potentially including varieties of predictive processing), and ‘integration’ theories (including workspace theories) (for these being prominent promising approaches, see Francken et al., 2021; for a critical review of a range of theories of consciousness see Doerig et al., 2020). Grouping the theories this way is particularly useful for contrasting different commitments to the distinction between conscious and unconscious mental states. A useful example to illustrate this distinction is the phenomenon of blindsight.

In blindsight (Weiskrantz, 1997; cf. Phillips, 2021), when the person looks at an X in the blind field, and responds accurately, there is a visual mental state about the shape of the stimulus. In that minimal sense, the person ‘sees’ the shape. We can characterize the mental state as the person sees that ‘the shape is X’. According to higher order theories (e.g. Rosenthal, 2005), what is missing in the blindsight case is a mental state that makes one aware of the seeing itself; i.e. one that asserts that ‘I am seeing that the shape is X’. This latter mental state is a mental state about a mental state: it is a higher order state (i.e. a type of metacognitive state). Conversely, a state about the world (e.g. “the shape is X”) is a first order state. According to higher order theory, for a mental state to be conscious there must be a relevant higher order state. By higher order theory, in the case of blindsight, it is precisely because the person earnestly and seriously denies seeing, that we are justified in regarding the seeing as unconscious (Brown et al., 2019; Lau & Rosenthal, 2011;).

The two main varieties of integration theory are global workspace theory and information integration theory. According to global workspace theory (GWT) (Baars, 1988;

Mashour et al., 2020), the brain consists of many unconscious processors which can only implement very specific functions. The output of each processor competes for access to a ‘global workspace’. When some output wins this competition, it is broadcast to many other processors (accounting for flexibility) and at the same time becomes conscious. Information integration theory (IIT) develops the view that conscious experiences are simultaneously integrated and differentiated (i.e. have high complexity). On IIT, a conscious brain needs high levels of both integration and differentiation in its dynamics; more specifically, IIT proposes that consciousness is identified with irreducible maxima of integrated information generated by a physical system (Oizumi et al., 2014). Tononi (2008) and Mediano et al. (2019) have conjectured relevant complexity measures that are high only when both integration and differentiation are high (for empirical support see Casali et al., 2013; for criticism see Merker et al., in press). Integration theories imply people should be able to use conscious information in flexible ways. In global workspace theory, broadcast of mental states makes these states accessible to a variety of ‘consumer’ cognitive systems, including those involved in decision making and cognitive control; in integrated information theory, a system with high integrated information means no two subsets of the system are informationally isolated. For example, a person with blindsight may be able to point to the place where a simple stimulus is, if forced to; but they cannot point to where the stimulus is not (Persaud & Cowey, 2008). The inability to flexibly or spontaneously use information is a key feature of unconscious states for integration theories.

There are other theories of conscious content. One of these, local recurrence theory, claims that a mental state becomes conscious when it elicits a wave of re-entrant neural processing (Lamme, 2010). Activation can be sustained by such loops. According to Lamme, these feedback loops (i.e. re-entrant processing) produce conscious experience of the content coded by the areas, even if the loop is local to a particular sensory area. This conscious

content is only verbally reportable when the re-entrant loop encompasses prefrontal and parietal brain regions.

Theoretical framing of the distinction between conscious and unconscious mental states applies to experimental data in the context of different kinds of measurement. Two types of measurement are key here: objective measures and subjective measures. Tests asking people to indicate what is in the world (for example, what word was displayed) are called objective measures. Tests asking people to indicate what mental state they are in (for example a confidence rating) are called subjective measures. Higher order theories motivate the use of subjective measures. Most integration theories do too; when the content of a mental state is widely integrated it will be available to higher order thoughts. Thus on both theories, showing knowledge on objective measures while showing lack of awareness of the knowledge on subjective measures indicates that the knowledge is unconscious. Integration theories motivate another test of knowledge being conscious, namely, one should have control over how the knowledge is used (Jacoby, 1991). So for example, if one was asked to pick the word that had NOT just been flashed, conscious seeing would enable one to pick the other brand name; unconscious seeing may draw one to the very word flashed, an impulse one may not know to oppose. On higher order theory, unconscious knowledge could in principle be used flexibly, just so long as it did not elicit appropriate higher order states. Such an example is considered below (Wan et al., 2008). Thus, the theories are not the same. If information were in the global workspace but not re-represented through the kind of higher-order representation needed by higher-order thought theory,, it would be conscious by global workspace theory yet unconscious by higher order theory.

Local recurrence theory motivates the use of objective measures; recurrent processing allows representations to last long enough to potentially elicit a discriminative response, but there is no reason why local recurrence must allow access to higher-order processing, and

thus no reason it should always allow verbal reports. (It has been argued that a discriminative response may be needed for perception per se to occur, whether conscious or unconscious; thus, being below the objective threshold rules out any perception at all, i.e. ruling out unconscious perception a priori; see debate here: Peters et al., 2017.)

The evidence for a distinction between conscious and unconscious mental states is presented, for memory, perception, learning, intentions, and attitudes. This evidence is considered within the context of higher order and global workspace theories of consciousness, since these currently provide the sharpest empirically tractable distinctions between conscious and unconscious mental states. The article does not explore in depth the neural basis for a distinction between conscious and unconscious states, but focuses on the behavioural issues that need to be addressed in any study of the conscious unconscious distinction.

Conscious versus Unconscious Memory

Construing memory as a persistent influence of a past event, can there be memory without the subject being aware of it as memory? Warrington and Weiskrantz (1974) inspired much work by describing people with amnesia who had poor recognition memory for a list of studied words, but who showed normal stem completion priming. Both people with and without amnesia showed priming (see Hayes et al., 2012). Because people with amnesia showed priming, there was some form of memory for the word presented earlier that on the surface does not depend on consciously remembering the word. However, a non-significant difference in priming between groups – as reported in this early study – is not evidence for no difference, a problem we will come back to.

It turns out that memories can present themselves as perceptions rather than memories. When people have heard sentences before, those same sentences sound clearer

when subsequently heard in noise. Jacoby et al. (1988) found that people rated the noise as less loud when listening to old rather than new sentences, even though the noise level was the same. Further, when told about this effect, and asked not to show it, people still thought the noise was quieter for old rather than new sentences. Jacoby argued that unconscious memory expresses itself as *fluency*, i.e. an ease of processing a stimulus, but the way the fluency is experienced depends on what the person is trying to do.

Familiarity with a claim may make the claim appear more credible. In a classic experiment, Hovland and Weiss (1951) asked people to read an essay on a controversial topic written by a credible or a non-credible source. When tested immediately, attitude change happened more for the credible source. But after a month's delay, the opinions of both groups were similar. Hovland and Weiss called the increasing influence of the low credibility source over time, the 'sleeping effect' (for boundary conditions on obtaining the effect see Pratkanis et al., 1988). The sleeping effect may be due to forgetting the source of the arguments, as conscious recollection of the details of that episode in the psychology lab fade (cf Begg et al., 1992). Nonetheless, when one is asked about the topic some time later, arguments may spring to mind. Now, with the source forgotten consciously, the arguments may seem like one's own. This same mechanism may be the basis of unconscious plagiarism: Declining recollection for an argument's source may lead one to plagiarize without knowing it is plagiarism. Marsh et al. (1997) asked participants in groups to brainstorm ways to (for example) improve the university. After a one-week delay, participants came back and were asked to generate four new ways to address the problem. Participants were specifically told not to repeat ideas from the previous session; nonetheless, about 20 per cent of their 'new' ideas could be shown to come from the initial session (cf Stark & Perfect, 2007).

The mere exposure effect refers to how exposure to a novel stimulus can lead people to like it more (Zajonc, 1980). For example, if novel shapes are flashed to people quickly,

then the more times they are flashed, the more people will like the shapes. A musical tune is liked more with repetition (up to a point; Szpunar et al., 2004). Similarly, exposing children to a new vegetable can substantially increase both their liking and subsequent intake of the food (Wardle et al., 2003). What role does conscious memory play in the mere exposure effect? The effect is stronger for supraliminal rather than subliminal exposure, but can still occur for subliminal stimuli (e.g. Kawakami & Yoshida, 2019). And although the mere exposure effect can be impaired by some types of brain damage that also impair conscious memory (e.g. to right temporal lobes), patients with severe global amnesia can have substantial and perhaps intact mere exposure effects (Marin-Garcia et al., 2013). In sum, the mere exposure effect may not require conscious recollection for the exposure episode.

What people hear while unconscious because of general anaesthesia can sometimes later influence them (Deeprrose & Andrade, 2006). This effect holds even when depth of anaesthesia has been ascertained. For example, Iselin-Chaves et al. (2005) found that people completed word stems at above baseline level for words presented during deep anaesthesia, even when patients has been instructed to avoid using the words presented during that period. Whether we can learn new information during sleep has remained unclear. Arzi et al. (2012; also Canales-Johnson et al., 2020), suggest that we can – at least when it comes to smell. They found that sleeping subjects were able to learn novel associations between particular odors and auditory tones, as shown by ‘selective sniffing’ the following day. Kouider et al. (2014) asked awake participants to classify words into different categories, while transitioning toward sleep. Strikingly, there were task-specific preparatory responses even after sleep onset, as revealed by so-called ‘lateralized readiness potentials’. However, even if some forms of simple conditioning and word processing can occur during sleep, that does not mean that sleep is a good way of learning in general. Wood et al. (1992) found no significant

implicit or explicit memory for words played during sleep, unless the person was immediately awakened.

An argument that could be used for the effect of previous exposure of a stimulus constituting an influence we are not aware of is that the effect would be irrational if produced consciously. Such an argument could be made for the phenomenon of social or behavioural priming. For example, Williams and Bargh (2008) found that if asked to hold a hot rather than cold drink briefly, participants later rated other people as being warmer in terms of personality. However, attempts to replicate directly any given social priming effect exactly have met with failure: for example, Lynott et al. (2014), with 861 participants, failed to replicate the warm coffee effect (see also e.g. Doyen et al., 2012; Pashler et al. 2012; Shanks et al. 2013). Social priming effects are smaller than many of the studies imply, and likely non-existent under many conditions. Under exactly what conditions such priming may occur remains to be established. Newell and Shanks (2014) point out that there is one form of irrational priming that is frequently replicated: the anchoring effect (first investigated by Tversky & Kahneman, 1974). People given a number for any reason, subsequently bias their estimates of a quantity (date, price, temperature, etc) towards the given number, even if it is explicitly indicated to be random.

In the cases discussed of unconscious memory, there is a past influence on judgments that the person is not aware of; but whether there is evidence of an unconscious memory per se depends on how broadly one wishes to construe what memory is. There is a modification of processing (that can be experienced consciously in e.g. perception), but is there any state that is unconscious? Clearer cases of possible unconscious mental states are considered next.

Conscious versus Unconscious Perception

Historically the most common method for creating subliminal perception has been a backward mask where for example, a string of letters masks a word; or one face masks another. The time from the beginning of the target to the beginning of the mask is called the stimulus onset asynchrony (SOA). The critical SOA to render a target subliminal is different for different people; it may be 15ms for one person and 200ms for another (e.g. Armstrong & Dienes, 2013). Two newer methods have been developed that allow the subliminal stimulus to be displayed for considerably longer than 100ms: continuous flash suppression (CFS) (Tsuchiya & Koch, 2005) and gaze contingent crowding (GCC) (Faivre et al., 2012). In CFS, a very salient mask is presented to one eye, namely a mixture of coloured blocks that are continuously changing. A static target is presented to the other eye. In this way, the target can be kept masked for many seconds before the target breaks through. In gaze contingent crowding (GCC), a peripheral target stimulus is surrounded by masks (i.e. ‘crowded’), which means the target is not consciously processed – so long as it remains in the periphery. In GCC, an eye-tracker is used such that as soon as the participant stops focusing straight ahead, the target disappears, and reappears when the participant focuses straight ahead again (i.e. contingent on the gaze). At the opposite end of the timing spectrum, there has been a return to purpose-built tachistoscopes that can present stimuli with a timing accuracy on the order of microseconds (Beauny et al., 2020).

Cheesman and Merikle (1984) first distinguished between the subjective threshold and objective threshold. People were shown a backward masked colour word (one of four). They indicated which word had been presented. They also gave a confidence rating, anything from 25 to 100 per cent after each block of trials. As the SOA was reduced, the subjective threshold was reached when the person gave a 25 per cent (chance level) confidence rating, i.e. they believed that were just guessing. The guessing criterion of unconscious knowledge claims unconscious knowledge is used when people perform above chance but believe they

are completely guessing. Crucially, at the subjective threshold, people were actually correct 66 per cent of the time. The SOA needed to be reduced still further to reach the objective threshold, the point where people actually classified the words at 25 per cent. After finding both thresholds for each participant, participants performed a Stroop task. Stroop priming was found below the subjective threshold; it was not significant below the objective threshold. Thus, on higher order and global workspace theories, subliminal perception was shown by Cheesman and Merikle.

Fisk and Haase (2005) showed that subjects can say “there was not a stimulus” on a detection task while being above chance at saying which stimulus was displayed on a discrimination task. This is a way of the discrimination being below a subjective threshold. However, detection sensitivity over all trials was higher than discrimination sensitivity, and when plotted against each other, both tended to go to zero together. Thus, Fisk and Haase argued that the subjective threshold was invalid as the pattern of findings can be explained in signal detection terms as simply reflecting what information is used to make which decision (detection vs discrimination). However, such an SDT approach still leaves open which information was conscious and which unconscious. For that, a theory of consciousness is needed.

A key issue to address when using subjective measures is the potential for ‘response bias’. When people in everyday life say ‘I guess that...’, they mean a range of things. They might even mean they are fairly confident. A key part of the solution is to devise ways of being more rigorous in the use of subjective measures. For example, ‘guess’ can be precisely defined as being no better than random, with examples (e.g. Cheesman & Merikle, 1984); people’s confidence may be elicited by gambling or wagering, so actual rewards are on the line (Dienes & Seth, 2010; Persaud & McLeod, 2008); confidence can be taken after every trial rather than after blocks of trials (Kolb & Braun, 1995), and so on. A further innovation

was to use labels devised to correspond most naturally to how participants themselves describe the clarity of their visual experience, leading to the influential Perceptual Awareness Scale (PAS), with levels of “no experience” (1), “brief glimpse” (but cannot say what it was) (2), “almost clear experience” (3), and “clear experience” (4) (Ramsøy & Overgaard, 2004). The difference between (2) and (3) marks the difference between being unaware of seeing the discrimination tested and being aware of seeing it (whether the difference between other scale points marks degree of consciousness has been a matter of debate; Michel, 2019). A challenge for the PAS is that it must be defined separately for each new experimental situation.

No amount of adjusting a measure in itself can validate it in the absence of theory (Chang, 2007). A further crucial argument for subjective measures measuring more than response bias, provided by Merikle (1992), is that the subjective threshold proves its worth when it separates qualitatively different types of perception. More precisely, if perception measured to be above and below the subjective threshold behaves like a theory of consciousness says it should, then the measure can pick itself up by its bootstraps. Merikle and Joordens (1997) presented one of two masked colour words followed by the incongruent colour 75 per cent of the time, with the task being to name the colour. Thus, if one could flexibly use the knowledge of what the word was, one should be able to predict that the colour would be the opposite, and hence be faster on the *incongruent* rather than the congruent trials; i.e. one would get a reverse Stroop effect. When the word was presented at above the subjective threshold, this was exactly what was found, a reverse Stroop effect. But for words below the subjective threshold, a normal Stroop effect was found. The results suggest that the subjective threshold has real consequences for perception and behaviour. Further, the consequences are as predicted by a theory of consciousness: If higher order thoughts are available only when information is in a global workspace, then perception above

rather than below the subjective threshold should involve more flexible use of the information. Subsequent studies have found there can be control based on congruency proportion for subliminal presentation (Pandero et al., 2015), but the dissociation between above and below subjective threshold remains (Lorentz et al., 2015). The evidence for theoretically motivated distinction between above and below subjective threshold remains limited however; a sobering thought given the number of decades that have elapsed since Merikle made his argument in 1992. Nonetheless, a survey of consciousness researchers at a prominent annual conference found a majority considered subjective measures the best way to check for conscious awareness, though a substantial minority endorsed objective measures (Francken et al., 2021).

A final method of addressing the bias problem in subjective measures is to attempt a bias-free measurement. For example, Rajananda et al. (2020) presented subjects with two intervals; in one there was a masked face displaying fear or happiness; in the other there was (unbeknownst to participants) a masked neutral face. For each face, subjects indicated whether the emotion was fear or happiness (the Type I discrimination, the measure of perception). In addition, participants were forced to choose which interval they were most confident about (the Type II discrimination, the measure of whether the perception was conscious). Because the Type II discrimination was forced choice, it can function as a bias free measure. They found that the Type I discrimination was above chance only when the Type II discrimination was as well. However, making a subjective measure bias free may be removing its function as a subjective measure. What is a participant to do when they believe they are equally likely to be accurate in either interval? They might answer a first order question, such as in which interval was the face most emotional? From a higher order thought perspective, this can be achieved with a purely (i.e. unconscious) first order mental state.

In general, it has been easy to obtain priming effects below a subjective threshold, and difficult below an objective threshold. However, using sensitive methods it has been possible to obtain evidence of processing seemingly below an objective threshold as well. For example, Faivre et al. (2012) masked faces using gaze contingent priming. The faces could have neutral happy or angry expressions. Participants were non-significantly different from chance at indicating which expression the faces were showing. The target was a Chinese character which was rated as pleasant or unpleasant. Smiling faces made the character look more pleasant than the other faces.

The objective threshold faces challenges which are routinely overlooked in its use. Being non-significantly different from chance is not evidence in itself that performance is at chance. To obtain evidence that performance is at chance there must be an estimate of what the performance would have been if it had accounted for subsequent priming (Dienes, 2015). Using one way of estimating an expected effect, Meyen et al. (in press) argued that many existing published demonstrations of subliminal effects at objective threshold did not show objective discrimination performance to be smaller than expected given the priming effect. Meyen et al. assume equal signal to noise ratios in the discrimination and priming tasks, which although simplistic, is a good way of putting the ball in the park of those arguing for the use of an objective threshold. Dienes recommended that the relation between priming and expected discrimination should be empirically calibrated; see Skora et al (2020) for an application. Finally, one potentially useful method for showing performance at objective threshold is to regress priming against discrimination and examine the slope and intercept; a zero slope or especially a non-zero intercept can be taken to indicate subliminal priming (Greenwald et al., 1995). Statistical methods are still being refined to take into account that poor reliability in the discrimination measure will artefactually flatten the slope and can itself

introduce a non-zero intercept (e.g. Malejka et al, in press, for use of Bayesian hierarchical modelling to deal with the problem of poor reliability).

In finding a threshold, a common logic is that one has found conditions under which the stimulus is never consciously perceived. But claims for subliminal perception would be undermined by any contamination by conscious perception. Jacoby et al. (1992) illustrated how to show subliminal perception one did not have to find conditions that reflect purely conscious or purely unconsciously processes; the Process Dissociation Procedure or PDP (see Jacoby, 2009). Jacoby et al. presented a masked word. In an *exclusion test*, the instruction was to ‘complete the stem with a word that comes to mind but not any you just saw displayed’. In the exclusion test, conscious perception of the word would lead one to make sure one did not use it; thus, any tendency to use the presented word at above baseline levels is due to unconscious perception. In exclusion, conscious and unconscious perception act in opposition. By contrast, in the *inclusion test*, the instruction was to ‘complete the stem with one of the words flashed or, if unable to do so, with the first word that comes to mind.’ In inclusion performance, conscious and unconscious perception act in concert; they would both lead one to complete the stem with the displayed word. Inclusion was greater than exclusion; so people had some control over the use of the information. Only conscious perception can allow this according to Jacoby et al. Performance on the exclusion task was greater than baseline showing a lack of control, which was taken as evidence for unconscious perception. Thus, for the same tasks and conditions, Jacoby et al. demonstrated a mix of conscious and unconscious perception.

In general, Jacoby’s method, PDP, takes flexible control to be the functional criterion for consciousness. Flexible control as a function of consciousness aligns with global workspace theory. In this particular case, the use of PDP also aligns with higher order theory. The exclusion task was another way of testing for the subjective threshold: The instruction

was to exclude a word if *you think you saw it*, i.e. if you have a relevant higher order thought (e.g. Sandberg et al., 2020), and so PDP can also face a response bias problem (Fu et al., 2008; Visser & Merikle, 1999).

Just as the PDP can allow conscious and unconscious processes to be dissociated in impure tasks, confidence ratings and other subjective measures can also be used to provide trial by trial classifications of whether perception or knowledge is conscious or unconscious, without assuming the task is always purely one or the other. For example, for a given SOA, the trials on which people say they are purely guessing can be taken as involving unconscious perception; and the trials on which people say they saw, or have some confidence, can be taken as conscious perception (e.g. Lau & Passingham, 2006). As long as subjective measures are taken trial by trial, there is no need to assume a task is performed purely consciously or purely unconsciously. However, when trials are post hoc selected by any variable with measurement error (including for example, reports of “guessing”), then a proportion of the trials classified as unconscious will in fact be conscious. That is, regression to the mean can apply to post hoc selection of subsets of trials or participants, leading to artefactual above chance performance on selected trials or subjects (Shanks, 2017). The problem can be addressed by Bayesian methods in the cases of post hoc subject selection (Leganes-Fonteneau et al., 2021) and both trial and subject selection (Skora et al., 2020).

Vision for Action

Goodale and Milner (1992) argued that the visual system contains two functionally different streams, the ventral stream (from early visual cortex to the temporal lobe) and the dorsal stream (from early visual cortex to the parietal lobe). The function of the ventral stream is to support the identification and recognition of objects; the function of the dorsal stream is to enable rapid interaction with the environment. The subliminal processing we

have considered so far has involved processing in the ventral stream, as it involved recognition of words and shapes. However, on this theory, unconscious processing may be more characteristic of the dorsal than ventral stream. Milner and Goodale found that patients with damage to the ventral stream, had difficulty verbally recognizing or describing the very same shapes that they could act on effectively. For example, DF when asked to reach for objects showed appropriate pre-shaping of her hand according to the size of the object; yet if she was asked to simply look and judge and size, she could not do it. Rossetti et al. (2017) reviewed evidence for disputing that the behavioural dissociation relies on specifically a dorsal vs ventral stream (cf. Ganel & Goodale, 2019, for a review defending the dorsal-ventral dissociation); however, the claim that action control can run independently of voluntary control is one Rossetti et al. support.

If action control can run independently of detailed conscious guidance, perceptual-motor skills may often involve subliminal perception. When people intercept a ballistic object, such as a cricket ball, people use a strategy involving the angle of gaze; i.e. the angle that the line of sight from the eye to the ball makes to the horizontal. If this angle is kept between 0° and 90° , the ball will be intercepted. Allowing the tangent of the angle to rise at a constant rate in fact ensures the angle will never exceed 90° . When people are filmed catching balls, they kept the tangent of the angle of gaze increasing at a constant rate. People's conscious theories of what they do are rather different, however. Reed et al. (2010) showed that people believed in rules that did not describe their behaviour and would in fact guarantee failure if they were to be followed. So here, people's unconscious knowledge, guiding the specific action of intercepting a ballistic object, was more accurate than their conscious beliefs about what they saw, based on general theories. Where there is a clear simple solution to an ecologically important problem (will a ballistic object intercept you?), it

is safer to keep performance shielded from flexible conscious knowledge (as the content of the latter will often guarantee performance failure).

Conscious versus Unconscious Learning

Implicit learning, a term coined by Reber (1967), is the acquisition of unconscious knowledge about the structure of the environment. The simplest form of structural learning is conditioning, where one learns about a simple association or reinforcement contingency. Obtaining unconscious conditioning has remained difficult (Lovibond & Shanks, 2002). Experimental attempts to induce unconscious conditioning typically use only a few salient stimuli; as soon as any unconscious conditioning has happened, attention may naturally be drawn to the key relationship, enabling conscious knowledge of the contingency. Even using non-salient incidental stimuli can still lead to fragile effects (Moran et al., 2020). A way to allow unconscious conditioning to show itself is to use stimuli that misdirect people's conscious hypotheses. People are susceptible to the gambler's fallacy, meaning that after a fair coin has produced a run of heads, people think it is 'about time' a tail came up.. Perruchet (1985) put the gambler's fallacy in opposition to associative learning. People received a puff of air to their eye after a tone 50 per cent of the time. The longer the run of a tone followed by a puff, the less people said they expected a puff on the next trial, in line with the gambler's fallacy – but the more strongly they blinked. Conversely, the longer the run of a tone alone, the more people said they expected a puff on the next trial – but the less strongly they blinked. That is, the conditioned response (blinking) and conscious expectancy went in opposite directions (see e.g. Destrebecqz et al., 2019, for evidence that the Perruchet effect as shown in a reaction time paradigm involves a dissociation between expectations and associative processes). The conditioned response therefore plausibly reflects unconscious structural learning.

One way to prevent conscious knowledge of a simple association is to present the stimuli subliminally. For example, Pessiglione et al. (2008) were the first to show instrumental learning with subliminal conditioned stimuli, apparently at objective threshold. The issues just reviewed concerning whether an objective threshold are reached are of course relevant; Skora et al. (2020) is a Stage 1 Registered Report replication of Pessiglione et al. using informed Bayesian methods to determine the presence of learning when there is genuine evidence for zero discrimination. The issue of having genuine evidence for no discrimination remains to be addressed generally. For an intriguing possible application of subliminal conditioning, Taschereau-Dumouchel et al. (2018) asked subjects to expand a green circle by thinking whatever they needed to, in order to make the circle expand. Unknown to the subject, the size of the green circle was determined by an online fMRI decoder that caused the circle to expand the better it could decode a feared animal. Over time, SCRs and BOLD response reduced specifically for images of the decoded animal. Yet when asked which of two animals the participants thought was being decoded, they were only 47% correct; and participants' free reports about their thoughts appeared to be of irrelevant things and activities such as singing. These results suggest that a subliminal form of exposure therapy may be possible. However, While although the results are promising, note the 95% confidence interval on discrimination performance went up to 64%. Further, the thought content reported, e.g. songs, may have been semantically associated to the feared object: We will come to the issue of "correlated hypotheses" below.

The artificial grammar learning paradigm introduced by Reber (1967) was inspired by natural language as an example of implicit learning. Reber asked people in an initial training phase to memorize or simply look at a number of letter strings generated by a finite state grammar, but without telling people the strings were structured. The grammar was complex enough not to be reconstructed explicitly in the time available. After 10 minutes of exposure,

people are informed that actually the strings they saw obeyed a complex set of rules; they are then asked to classify new strings as rule governed or not in the subsequent test phase. People could classify the new strings at about 60–70 per cent correct, depending on the grammar used, even though in free report afterwards they found it difficult to say what the rules were. However, free report is often not a very sensitive test of conscious knowledge (contrast Paciorek & Williams, 2015). People may miss out saying something simply because they momentarily forgot it; or because they are not quite confident enough to risk saying something that might be wrong; or, if people think the experimenter is interested in one sort of answer, they may indicate ignorance (for example, if they believe the experimenter is interested in rules, but the participant used analogy to memorized exemplars, Brooks, 1978).

There are other issues with the basic artificial learning grammar paradigm. Dulany et al. (1984) argued that participants might not learn the finite state grammar as such but “correlated hypotheses”; that is, many small rules concerning the allowability of chunks of letters in given positions, each rule of imperfect validity. Using an objective measure of knowledge, participants underlined valid chunks or crossed out invalid ones in the to-be-classified strings; the validity of these rules predicted classification performance. Dulany et al. argued the procedure entailed participants reporting conscious rules; but using a forced-choice response does not as such constitute report. Participants may have been using unconscious knowledge with the phenomenology of guessing - or using intuition (Reber et al., 1985).

One solution to these issues is to take confidence ratings every trial, then the problems just mentioned with either free report or using objective measures are side stepped. If people classify above chance when they believe they are purely guessing the guessing criterion of unconscious knowledge is satisfied; and if there is no relation between confidence and accuracy, the zero correlation criterion is also satisfied. Satisfying these two criteria can

indicate the presence and amount of unconscious knowledge. As in the case of subliminal perception, subjective measures can be criticized as dependent on a participant's decisions as to how to use a confidence scale, a point we will return to.

When claiming a mental state is conscious or unconscious, it is important to specify what contents define that mental state. When confidence ratings are used, what knowledge exactly can be claimed to be conscious or unconscious? In the training phase, people acquire knowledge of the structure of the strings. This knowledge may consist in part of, for example, bigrams and other chunks, or what letters can start a string. Call this structural knowledge. In the test phase, the participant makes a judgement that, for example, this string is grammatical. The content of the judgement is something one knows about; call this judgement knowledge. The structural knowledge is the knowledge that enabled the judgement, but the two have different contents (e.g. 'M cannot start a string' vs. 'MTTVT does not have the right structure'). Confidence ratings are indications of confidence in the judgement. Thus, they reflect higher order thoughts about specifically judgement knowledge. That is, confidence ratings do not directly measure the conscious status of structural knowledge. If structural knowledge is unconscious, judgement knowledge could be conscious (judgement apparently based on intuition) or unconscious (judgment apparently based on guessing). In the first case, criteria based on confidence ratings (such as the zero correlation and guessing criteria) might show all knowledge is conscious – because those criteria only assess judgement knowledge.

Dienes and Scott (2005; see also e.g. Rebuschat, 2013) asked participants in the test phase of an artificial grammar learning experiment to indicate the basis of the judgement. Participants could indicate guess, intuition, rules or memory. The first two attributions (guess and intuition) reflect unconscious structural knowledge and the second two (rules and memory) reflect conscious structural knowledge. For all attributions, people classified the test strings as rule governed significantly above chance, indicating both conscious and

unconscious judgement and unconscious structural knowledge. But, taking a sceptical stance, we can ask why should we believe that when participants give these structural knowledge attributions they are actually picking out real differences in their mind? A scientific measure can only ever prove itself by showing its sensitivity to theoretical predictions. Evidence is emerging (but still limited) that the attributions show such sensitivity. For example, the accuracy of unconscious structural knowledge was unaffected by whether people were forewarned about rules in the training phase or by whether people were consciously distracted by a demanding secondary task; conversely, the accuracy of conscious structural knowledge was harmed by being distracted while searching for rules. An important question for any conjectured measure is under what conditions the measure picks out real differences, and this remains a matter for future research. A potentially useful line of enquiry may be how unconscious vs conscious structural knowledge is constrained in what can be learned (cf. Chen et al., 2011)

Jacoby (1991) argued that unconscious knowledge was inflexible and so not responsive to one's intentions whether to use it or not. However, Wan et al. (2008) and Norman et al. (2019) exposed people to sequences from two grammars. The test phase showed sequences from both grammars but participants were asked to pick from just one of the grammars. Can people strategically choose which grammar to use? The answer is that people can, and for every type of attribution. That is, whether or not judgement or structural knowledge is conscious or not, a person can decide to use one grammar and effectively use it (just as a bilingual can control which language they use).

Implicit learning involves not only coming to judge accurately how well-formed things feel, but also being able to make fast responses in the right situation. In an early study, Nissen and Bullmer (1987) designed a serial reaction time (SRT) task, in which people sped up in indicating what stimulus was presented when exposed to a structured sequence.

Destrebecqz and Cleeremans (2001) trained people on an SRT task, then asked people to generate a sequence. In the inclusion condition they were asked to generate the same sequence they were trained on; in the exclusion condition they were asked to generate a sequence different from the one trained on. Crucially, on the exclusion task, people generated fragments from the sequence at above chance levels – suggesting that knowledge of sequence fragments was unconscious. Fu et al. (2008) indicated the conditions under which above chance exclusion performance can be obtained. What knowledge does this procedure measure the conscious status of? The answer is judgement knowledge (i.e. the knowledge that an item is in the right place in a sequence). If one confidently judges that a production of an item is grammatical one can include or exclude it. But one does not have to know why a production is grammatical. However, given that at least some of the judgement knowledge was unconscious, we can infer some structural knowledge was too (see Fu et al., 2010). Vékony et al. (2020) showed an ASRT task (Alternating Serial Reaction Time task in which every other trial is random) was resistant to a secondary task (counting stimuli) at testing, with evidence of above chance exclusion performance – consistent with the application of unconscious structural knowledge being resistant to effects of diminished executive capacity.

Conscious versus Unconscious Intentions: Volition and Hypnotic Response

Can a person intend to perform an action, but think that they had no such intention? Consider hypnotic actions, which can be as complex as one wishes to suggest, given only that the voluntary version of the action is within the normal repertoire of the subject (see e.g. Palfi et al., 2021): From an arm rising by itself to automatic writing. Spanos (e.g. 1986) and Hilgard (e.g. 1977) provided evidence for the strategic and intentional nature of hypnotic response. In one example, Spanos et al. (1987) gave highly hypnotizable subjects ('highs') the post-hypnotic suggestion that they would rub their eyebrow whenever they heard the

word ‘experiment’, over the next 48 hours. Highly hypnotisable subjects did not respond to the suggestion when approached on campus by a confederate. That does not mean highs were faking when they did respond; it means a response that genuinely felt involuntary was constructed to occur so as to fit in with the subject’s overall goals. In one sense, then, the person responding hypnotically is still in charge of themselves, even if it feels to them like they are not. Dienes and Perner (2007) argued that hypnotic response involves acting on executive intentions (hence acting in a way consistent with one’s goals) while formulating the inaccurate higher order thought that one does not have the intention to act (that is, the intention is unconscious), the theory they called the cold control theory.

Phenomenological control is the capacity to control phenomenology in any appropriate context (for example, by using cold control, Dienes et al, in press). Thus, participants may create the illusion of a rubber hand being theirs if demand characteristics indicate this is the phenomenology they need to create (Lush et al, 2020). Unconscious intentions – giving rise to phenomenological control - in responding to demand characteristics is a methodological issue that seems to have been overlooked in many paradigms.

Conscious versus Unconscious Attitudes

An ‘attitude’ in the meaning used here is a positive or negative evaluation. Can liking, or attitude, itself be unconscious? Some evidence for this was provided by Adam et al., (1996) found homophobic men were considerably more aroused by male homosexual clips than the non-homophobic men. One plausible theory is that homophobes have unconscious positive feelings about gay sex. Or maybe homophobes are reacting to their purely conscious feelings, feelings they just didn't want to explicitly inform people about.

To obtain evidence for an unconscious attitude we need a way of measuring attitude that is not based on self-report; indeed, a measure that can contradict self-report in order to show that the attitude is unconscious. The implicit association test (IAT) provides one possibility (Banaji and Greenwald, 2013). On this test, words are classified by pressing one button for pleasant and the other for unpleasant words. In addition, the participant makes another classification using the same keys – for example, names of black people with one button and white people with the other. The question is, is a person faster when white rather than black is paired with pleasant rather than unpleasant? If so, that may indicate prejudice for white people and against black people. An initial key finding was that a greater proportion of people were measured as prejudiced by the IAT task than by their explicit reports for many types of prejudice (Banaji & Greenwald, 2013). But does this mean that people's negative attitudes to certain groups were unconscious? According to Gawronski and Bodenhausen (2006), attitudes as measured by the IAT reflect spontaneous conscious reactions, regardless of whether one thinks that these reactions are valid or invalid. Conversely, self-reported attitudes result from working out one's beliefs based on consistency with any other relevant belief (e.g. that all people should be treated equally, regardless of race). Thus, a person might decide that an initial reaction to a person of a different race is inconsistent with their general values and commitments – and thus give different explicit reports than their IAT indicates. Hahn et al. (2014) directly tested this idea by asking people to predict their IAT scores for a range of groups they could be prejudiced to; their predictions correlated about 0.6 with their actual scores. That is, people seemed quite aware of what their automatic attitudes were. This occurred despite explicit reports of prejudice having a small correlation with the IAT – that is, one's considered opinion need not be the same as an immediate reaction. The IAT may measure automatic evaluative associations, but that doesn't mean the attitudes are

unconscious. Whether the IAT, or similar measures, can pick up on unconscious attitudes in some contexts remains a tantalizing topic for future research.

Conclusion

This chapter has considered how we can define the notion of consciousness. We did this by first outlining two major theories of consciousness: higher order theories and integration theories. We highlighted these theories because, among the current crop of theories, they most naturally address the distinction between conscious and unconscious mental states. These theories motivate subjective measures - with which it is relatively easy to demonstrate unconscious mental states: memory, perception, knowledge and intentions. The main criticism of subjective measures is that they may reflect response bias. One way forward is to use theories of conscious vs unconscious mental states to determine if the measures behave as predicted. A start has been made in this direction. If one adopted local recurrency theory, most results reported here would not in fact be evidence for unconscious mental states (as it is harder to get evidence for unconscious states using objective measures). Recently the field has focused on the neural correlates of consciousness (NCC) and often in that context taking the measures of the conscious status of mental states for granted (for reviews see Frith, 2021; *Neuropsychologia*, vol 128; Williams, 2020) . A promising approach here is to connect theories more closely to candidate NCCs by elaborating candidate theories in terms of underlying neurocomputational mechanisms. Doing so would prevent the conscious status of mental states from being taken for granted, instead opening it up to theoretically-informed empirical investigation. Altogether, much remains to be done in constructing and validating measures of unconscious mental states.

Further reading.

- Greenwald, A. G., & Banerjee, M. R. (2017). The Implicit Revolution: Reconceiving the Relation Between Conscious and Unconscious. *American Psychologist*, 72, 861– 871.
- Peters, M. A. K., Kentridge, R. W., Phillips, I., Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate, *Neuroscience of Consciousness*, Volume 2017, Issue 1, nix015, <https://doi.org/10.1093/nc/nix015>
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63 (3), 595-626
- Seth, A. K., & Bayne, T. (in press). Theories of consciousness. *Nature Reviews Neuroscience*,
- Steven, S. J., Kirsch I, Terhune, D. B., & Green, J. P. (2020). Myths and Misconceptions About Hypnosis and Suggestion: Separating Fact and Fiction. *Applied Cognitive Psychology*, <https://doi.org/10.1002/acp.3730>

References

- Adams, H.E., Wright Jr, L. W., & Lohr, B. A. (1996). Is homophobia associated with homosexual arousal? *Journal of Abnormal Psychology*, 105(3), 440–445.
- Armstrong, A. M., & Dienes, Z. (2013). Subliminal Understanding of Negation: Unconscious Control by Subliminal Processing of Word Pairs. *Consciousness & Cognition*, 22 (3), 1022–1040.
- Arzi, A., Shedlesky, L., Ben-Shaul, M., Nasser, K., Oksenberg, A., Hairston, L. S., & Sobel, N. (2012). Humans can learn new information during sleep. *Nature Neuroscience*, 15, 1460–1465

- Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge University Press: Cambridge.
- Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. Delacorte Press: New York.
- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effect of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- Beauny, A., De Heering, A., Muñoz Moldes, S., Martin, J. R., De Beir, A., & Cleeremans, A. (2020). Unconscious categorization of sub-millisecond complex images. *PloS one*, doi.org/10.1371/journal.pone.0236467
- Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121, 446–458.
- Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization* (pp.169-211). Hillsdale, N.J.: Erlbaum.
- Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Science*, Sep;23(9):754-768. doi: 10.1016/j.tics.2019.06.009.
- Canales-Johnson, A., Merlo, E., Bekinschtein, T. A., Arzi, A. (2020). Neural Dynamics of Associative Learning during Human Sleep. *Cerebral Cortex*, 30, 1708–1715.
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S. et al. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and

- Behavior. *Science Translational Medicine*, 5, 198ra105; DOI: 10.1126/scitranslmed.3006294
- Chang, H. (2007). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.
- Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, 36, 387–395.
- Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., & Dienes, Z. (2011). Unconscious Structural Knowledge of Form-meaning Connections. *Consciousness & Cognition*, 20, 1751–1760.
- Deepröse, C., & Andrade, J. (2006). Is priming during anesthesia unconscious? *Consciousness and Cognition*, 15(1), 1–23.
- 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., Vande Velde, M., San Anton, E., Cleeremans, A., & Bertels, J. (2019). Saving the Perruchet effect: A role for the strength of the association in associative learning. *Quarterly Journal of Experimental Psychology*, 72, 1379–1386.
- Dienes, Z., Lush, L., Palfi, B., Rooseboom, W., Scott, R., Parris, B., Seth, A., & Lovell, M. (in press). Phenomenological control as cold control: Hypnosis and beyond. *Psychology of Consciousness: Theory, Research, and Practice*.
- Dienes, Z., & Perner, J. (2007). The cold control theory of hypnosis. In G. Jamieson (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective*. Oxford University Press, pp 293–314.

- Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, 69, 338–351.
- Doerig, A., Schurger, A. & Herzog, M. H. (2021) Hard criteria for empirical theories of consciousness, *Cognitive Neuroscience*, 12:2, 41-62, DOI: 10.1080/17588928.2020.1772214
- Doyen, S., Klein, O., Pichon, C-L, & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind? *PLoS ONE* 7(1): e29081. doi:10.1371/journal.pone.0029081
- Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, 113(4), 541–555
- Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious Influences from Emotional Faces: A Comparison of Visual Crowding, Masking, and Continuous Flash Suppression. *Frontiers in Psychology* 3:129.
- Gary D. Fisk, G. D., & Haase, S. J. (2005). Unconscious Perception or Not? An Evaluation of Detection and Discrimination as Indicators of Awareness. *The American Journal of Psychology*, 118, 183-212
- Francken, J., Beerendonk, L., Molenaar, D., Fahrenfort, J. J., Kiverstein, J., Seth, A., & van Gaal, S. (2021). An academic survey on theoretical foundations, common assumptions and the current state of the field of consciousness science. <https://doi.org/10.31234/osf.io/8mbsk>
- Frith, C. D. (2021). The neural basis of consciousness. *Psychological Medicine*, 51, 550 – 562. DOI: <https://doi.org/10.1017/S0033291719002204>

- Fu, Q., Dienes, Z., & Fu, X. (2010). Can unconscious knowledge allow control in sequence learning? *Consciousness & Cognition*, *19*, 462–475.
- Fu, Q., Fu, X., & Dienes, Z. (2008). Implicit sequence learning and conscious awareness. *Consciousness and Cognition*, *17*, 185–202.
- Ganel, T., & Goodale, M. A. (2019). Still holding after all these years: An action-perception dissociation in patient DF. *Neuropsychologia*, *128*, 249-254,
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.
- Goodale, M. A.; Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15* (1), 20–25.
- Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible(“subliminal”) stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General*, *124*(1), 22–42
- Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*, 1369–1392.
- Hayes, S. M., Fortier, C. B., Levine, A., Milberg, W. P., & McGlinchey, R. (2012). Implicit Memory in Korsakoff’s Syndrome: A Review of Procedural Learning and Priming Studies. *Neuropsychology Review*, *22* (2), 132–153
- Hilgard, E. R. (1977). *Divided consciousness: Multiple controls in human thought and action*. New York: Wiley-Interscience.
- Hovland, C. I., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly*, *15* (4), 635–650

- Iselin-Chaves, I. A., Willems, S. J., Jermann, F. C., Forster, A., Adam, S., & Van der Linden, M. (2005). Investigation of Implicit Memory during Isoflurane Anesthesia for Elective Surgery Using the Process Dissociation Procedure. *Anesthesiology*, *103*, 925–933
- Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jacoby, L. L. (2009). Memory, process-dissociation procedure. In Bayne Tim, Cleeremans Axel & Wilken Patrick (Eds.), *The Oxford Companion to Consciousness*. Oxford University Press (pp 430—432).
- Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 240–247.
- Jacoby, L. L., Toth, J. P., Lindsay, D. S., & Debnar, J. A. (1992). Lectures for a Layperson: Methods for Revealing Unconscious Processes. In R. Bornstein & T. S. Pittman Eds), *Perception without Awareness: Cognitive, Clinical, and Social Perspectives*. New York: The Guildford Press.
- Kawakami, N. & Yoshida, F. (2019). Subliminal versus supraliminal mere exposure effects: Comparing explicit and implicit attitudes. *Psychology of Consciousness: Theory, Research, and Practice*, *6*, 279-291.
- Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature*, *377*, 336 – 338.
- Kouider, S., Andrillon, T., Barbosa, L. S., Goupil, L., & Bekinschtein, T. A. (2014). Inducing Task-Relevant Responses to Speech in the Sleeping Brain. *Current Biology*, *24*(18), 2208–2214.
- Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–240.

- Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 18763–8.
- Lau, H. C., & Rosenthal, D. (2011). Empirical Support for Higher-Order Theories of Conscious Awareness. *Trends in Cognitive Sciences*, 15, 365–373.
- Leganes-Fonteneau, M., Scott, R., Duka, T., & Dienes, Z. (2021). Avoiding pitfalls: Bayes factors can be a reliable tool for post hoc data selection in implicit learning. *Psychonomic Bulletin & Review*, <https://doi.org/10.3758/s13423-021-01901-4>
- Lorentz, E., Gould, L., Mickleborough, M., Ekstrand, C., Boyer, M., Cheesman, J., & Borowsky, R. (2015). All in one fell Stroop: Examining consciousness thresholds with a multiple response paradigm. *Psychology of Consciousness: Theory, Research, and Practice*, 2(2), 111–133.
- Lovibond, P. F. & Shanks, D. R. (2002) The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 3–26.
- Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J., & Dienes, Z. (2020). Phenomenological control: response to imaginative suggestion predicts measures of mirror touch synaesthesia, vicarious pain and the rubber hand illusion. *Nature Communications*, 11,4853 (2020). <https://doi.org/10.1038/s41467-020-18591-6>
- Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., & O'Brien, K. (2014) Replication of 'Experiencing Physical Warmth Promotes Interpersonal Warmth' by Williams and Bargh (2008). *Social Psychology*, 45 (3), 216–222.

Malejka, S., Vadillo, M. A., Dienes, Z., Shanks, D. R. (in press). Correlation analysis to investigate unconscious mental processes: A critical appraisal and mini-tutorial.

Cognition

Marin-Garcia El, Ruiz-Vargas JM, Kapur N. (2013) Mere exposure effect can be elicited in transient global amnesia. *Journal of Clinical and Experimental Neuropsychology*, 35(10), 1007–1014.

Marsh, R. L., Landau, J. D. and Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 23, 886–897. Mediano, P., Seth, A., & Barrett, A. (2018). Measuring Integrated Information: Comparison of Candidate Measures in Theory and Simulation. *Entropy*, 21(1), 17. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/e21010017>

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, 105, 776-798.

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

Merikle, P. M., & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition*, 6, 219–236.

Merker, B., Williford, K., & Rudrauf, D. (2021). The Integrated Information Theory of consciousness: A case of mistaken identity. *Behavioral and Brain Sciences*, 1-72. doi:10.1017/S0140525X21000881

Meyen, S., Zerweck, I.A., Amado, C., von Luxburg, U., & Franz, V. H. (in press).

Advancing Research on Unconscious Priming: When can Scientists Claim an Indirect

- Task Advantage? *Journal of Experimental Psychology: General*
<https://doi.org/10.1037/xge0001065>
- Michel, M. (2019). The Mismeasure of Consciousness: A Problem of Coordination for the Perceptual Awareness Scale. *Philosophy of Science*, 86, 1239–1249.
- Moran, T., Hughes, S., Hussey, I., ... & De Houwer, J. (2021). Incidental Attitude Formation via the Surveillance Task: A Preregistered Replication of the Olson and Fazio (2001) Study. *Psychological Science*, 1, 20-131.
- Newell, B. R., & Shanks, D. R. (2014). Prime numbers: Anchoring and its implications for theories of behavior priming. *Social Cognition*, 32, 88–108.
- Nissen, M. J., & Bullermer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, 19, 1–32.
- Norman, E., Scott, R. B., Price, M. C., Jones, E., & Dienes, Z. (2019). Can unconscious structural knowledge be strategically controlled? In Cleeremans, A., Allakhverdov, V., & Kuvaldina, M. (Eds), *Implicit learning: 50 years on*. Routledge (pp 159-173).
- Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Comput Biol*. May 8;10(5):e1003588. doi: 10.1371/journal.pcbi.1003588.
- Paciorek, A., & Williams, J. (2015). Implicit learning of semantic preferences of verbs. *Studies in Second Language Acquisition*, 37(2), 359-382.
- Palfi, B., Parris, B. A., McLatchie, N., Kekecs, Z., & Dienes, Z. (2021). Can unconscious intentions be more effective than conscious intentions? Test of the role of metacognition in hypnotic response. Registered Report. *Cortex*, 135, 219-239.

- Panadero, A., Castellanos, M. C., Tudela, P. (2015). Unconscious context-specific proportion congruency effect in a stroop-like task. *Consciousness and Cognition*, 31, 35-45.
- Pashler, H., Coburn, N., and Harris, C. (2012). Priming of social distance? Failure to replicate effects on social and food judgements. *PloS ONE*, 7(8):, e42510
- Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Pavlovian Journal of Biological Science*, 20(4), 163–170.
- Persaud, N., & Cowey, A. (2008). Blindsight is unlike normal conscious vision: Evidence from an exclusion task. *Consciousness and Cognition*, 17, 1050-1055.
- Persaud, N., & Mcleod, P. (2008). Wagering Demonstrates Subconscious Processing in a Binary Exclusion Task. *Consciousness and Cognition*, 17 (3), 565–575.
- Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal Instrumental Conditioning Demonstrated in the Human Brain. *Neuron*, 59, 561–567.
- Peters, M. A. K., Kentridge, R. W., Phillips, I., Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate, *Neuroscience of Consciousness*, Volume 2017, Issue 1, nix015, <https://doi.org/10.1093/nc/nix015>
- Phillips, I. (2021). Blindsight is qualitatively degraded conscious vision. *Psychological Review*, 128(3), 558–584. <https://doi.org/10.1037/rev0000254>
- Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper effect. *Journal of Personality and Social Psychology*, 54(2), 203–218.

- Rajananda, S., Zhu, J., & Peters, M. A. K. (2020). Normal observers show no evidence for blindsight in facial emotion perception, *Neuroscience of Consciousness*, Volume 2020, Issue 1, niaa023,
- Ramsøy, T.Z., Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, 3, 1–23.
- Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 317–327.
- Reber, A. S., Allen, R., & Regan, S. (1985). Syntactical learning and judgment, still unconscious and still abstract: Comment on Dulany, Carlson, and Dewey. *Journal of Experimental Psychology: General*, 114(1), 17–24
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*. 63, 595–626.
- Rebuschat, P. & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Studies in Second Language Acquisition*, 33, 829–856.
- Reed, N., McLeod, P., & Dienes, Z. (2010). Implicit knowledge and motor skill: What people who know how to catch don't know. *Consciousness & Cognition*, 19, 63–76.
- Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press
- Rossetti, Y., Pisella, L., McIntosh, R. D. (2017). Rise and fall of the two visual systems theory. *Annals of Physical and Rehabilitation Medicine*, 60 (3), 130-140.
- Sandberg, K., Del Pin, S. H., Overgaard, M., & Bibby, B. M. (2020, May 16). A window of subliminal perception. PsyArXiv, <https://doi.org/10.31234/osf.io/6v2pz>

- Shanks, D.R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychon Bull Rev* 24, 752–775.
<https://doi.org/10.3758/s13423-016-1170-y>
- Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., et al. (2013) Priming Intelligent Behavior: An Elusive Phenomenon. *PLoS ONE* 8(4): e56515.
- Skora, L., Livermore, J. J. A., Dienes, Z., Seth, A., & Scott, R. B. (2020). Feasibility of Unconscious Instrumental Conditioning: A Registered Replication. *Cortex*,
<https://doi.org/10.31234/osf.io/p9dgn>
- Spanos, N. (1986). Hypnotic behaviour: a social–psychological interpretation of amnesia, analgesia, and ‘trance logic.’ *Behavioural and Brain Sciences*, 9, 449–502.
- Spanos, N. P., Menary, E., Brett, P. J., Cross, W., & Ahmed, Q. (1987). Failure of posthypnotic responding to occur outside the experimental setting. *Journal of Abnormal Psychology*, 96(1), 52–57.
- Stark, L.-J., & Perfect, T. J. (2007). Whose idea was that? Source monitoring for idea ownership following elaboration. *Memory*, 15 (7), 776–783.
- Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and Memory for Musical Stimuli as a Function of Exposure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30(2), 370–381
- Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J. D., Kawato, M., Lau, H. (2018). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences*, 115 (13) 3470-3475;
- Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biological Bulletin*, 215, 216–242.

- Tsuchiya, N., & Koch, C. (2005) Continuous Flash Suppression Reduces Negative Afterimages. *Nature Neuroscience*, 8, 1096–1101.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- Vékony T, Török L, Pedraza F, Schipper K, Pleche C, Tóth L, et al. (2020) Retrieval of a well-established skill is resistant to distraction: Evidence from an implicit probabilistic sequence learning task. *PLoS ONE* 15(12): e0243541.
<https://doi.org/10.1371/journal.pone.0243541>
- Visser, T. A. W., & Merikle, P. M. (1999). Conscious and Unconscious Processes: The Effects of Motivation. *Consciousness and Cognition*, 8, 94-113.
- Wan, L. L., Dienes, Z., & Fu, X. L. (2008) Intentional control based on familiarity in artificial grammar learning. *Consciousness and Cognition*, 17, 1209–1218.
- Wardle, J., Herrera, M.-L., Cooke, L., & Gibson, E. L. (2003). Modifying children's food preferences: the effects of exposure and reward on acceptance of an unfamiliar vegetable. *European Journal of Clinical Nutrition*, 57, 341–348
- Warrington, E.K. and Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, 12, 419–428.
- Weiskrantz, L. (1997). *Consciousness Lost and Found: A Neuropsychological Exploration*. Oxford: Oxford University Press.
- Williams, J. N. (2020). The neuroscience of implicit learning. *Language Learning*, 70, 255-307.
- Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, 322, 606–607.

Wood, J. M., Bootzin, R. R., Kihlstrom, J. F., & Schacter, D. L. (1992). Implicit and Explicit Memory for Verbal Information Presented during Sleep. *Psychological Science*, 3 (4), 236–239.

Zajonc, R. B. (1980). Feeling and thinking: preferences need no inferences. *American Psychologist*, 35(2), 151–175.