Journal of Experimental Psychology: General
2017, Vol. 146, No. 12, 1705–1721

0096-3445/17/$12.00 http://dx.doi.org/10.1037/xge0000371

Unconscious Conditioning: Demonstration of Existence and Difference From Conscious Conditioning

Anthony G. Greenwald
University of Washington

Jan De Houwer
Ghent University

Unpronounceable strings of 4 consonants (conditioned stimuli: CSs) were consistently followed by familiar words belonging to 1 of 2 opposed semantic categories (unconditioned stimuli: USs). Conditioning, in the form of greater accuracy in rapidly classifying USs into their categories, was found when visually imperceptible (to most subjects) CSs occupied ≥58 ms of a 75-ms CS–US interval. When clearly visible CSs were presented in a 375 ms CS–US interval, conditioning was strongly correlated with measures of contingency awareness, and did not occur in the absence of that awareness. These experiments delineated 2 forms of conditioning: *Unconscious conditioning* occurred with a brief CS–US interval, with an effectively masked conditioned stimulus (CS), and with no reportable knowledge of the contingent CS–US relation. *Conscious conditioning* occurred with a substantially longer CS–US interval, a perceptible CS, and with subjects' reportable knowledge of the contingent CS–US relation.

Keywords: conditioning, unconscious cognition, response window method, visual masking, adversarial collaboration

Supplemental materials: http://dx.doi.org/10.1037/xge0000371.supp

Conditioning involves a change in behavior due to the pairing of a conditioned stimulus (CS) with an unconditioned stimulus (US). Because conditioning has been demonstrated in organisms with nervous systems much less complex than those of the original mammalian subjects of conditioning research (e.g., Thompson & McConnell, 1955), occurrence of conditioning in the absence of conscious cognition has not been at issue—for nonhuman subjects. The possibility of unconscious conditioning in human subjects was assumed by early researchers on human conditioning (Razran, 1955). In more recent years, however, this possibility has been forcefully contested (Lovibond & Shanks, 2002; Shanks, 2010; Vadillo, Konstantinidis, & Shanks, 2016; Weidemann, Satkunarajah, & Lovibond, 2016).

Rejection of unconscious human conditioning has recently taken three forms: First, claimed conditioning effects have been asserted to be nonassociative artifacts of procedures other than CS–US pairings, such as sensitization either to the CS or the US stimuli (Lovibond & Shanks, 2002). Second, published unconscious conditioning findings have been reported to fail in independent replications (e.g., Hendrickx, De Houwer, Baeyens, Eelen, & Van Avermaet, 1997). And third, claimed evidence for conditioning in the absence of contingency awareness has been argued to be an artifactual consequence of methodologically or conceptually weak procedures in measuring contingency awareness (Lovibond & Shanks, 2002; Vadillo et al., 2016; Shanks & St. John, 1994).[1]

This article reports findings that overcome these three empirical critiques. Additionally, the present findings that show unconscious conditioning are accompanied by findings of *conscious* conditioning effects that have properties clearly distinct from those found here for unconscious conditioning.

Studies of human eyeblink conditioning (Clark & Squire, 1998; Clark, Manns, & Squire, 2002) have described distinct conscious ("declarative") and nonconscious ("nondeclarative") conditioning processes that are assumed, respectively, to involve propositional and associative mental processes. Critics of such two-process views of human conditioning (Lovibond & Shanks, 2002; Mitch-

Anthony G. Greenwald, Department of Psychology, University of Washington; Jan De Houwer, Department of Experimental-Clinical and Health Psychology, Ghent University.

Preparation of this article was made possible by the Implicit Cognition Research Fund at University of Washington, by National Institute of Mental Health Grant MH–01533, by the Ghent University Grant BOF16/MET_V/002, and by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33). We thank Larry Squire, David Shanks, Peter Lovibond, and Allan Wagner for comments on a preliminary draft, and Maarten De Schryver for data collection at Ghent University.

Portions of the research have been presented at colloquia and small conferences. No portions of this research have been previously published.

A complete archive of the computer programs for data collection, raw data for all experiments, data analysis scripts needed to reproduce analyses, and outputs of data analyses is available at Open Science Framework: https://osf.io/ydeza/

Correspondence concerning this article should be addressed to Anthony G. Greenwald, Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195-1525. E-mail: agg@u.washington.edu

[1] Similar criticisms have been offered also for other unconscious learning paradigms—including artificial grammar learning (Dulany, Carlson, & Dewey, 1984; A. S. Reber, 1967; R. Reber & Perruchet, 2003), operant learning (Dulany, 1961; Krasner, 1958; Mastropasqua & Turatto, 2015; Van Dessel, De Houwer, Roets, & Gast, 2016), and perceptual learning (Seitz & Watanabe, 2005, 2008).

ell, De Houwer, & Lovibond, 2009) claim that successful conditioning requires both a propositional representation of the CS–US contingency and conscious awareness of that contingency. A recent review (Mitchell et al., 2009) appropriately observed that "a demonstration of unaware conditioning would be highly damaging to the propositional approach, and would provide strong evidence for a second (automatic) learning mechanism."

Seeking to resolve debates about theoretical understanding of conditioning, this research applied methods that helped to resolve an earlier contentious debate about the necessity for conscious cognition in unconscious ("subliminal") priming. Priming most often involves the effect of a first (prime) stimulus on the response to a second (target) stimulus. Conversely, conditioning most often involves the effect of a second (US) stimulus on the response to a first (CS) stimulus. In the early 1980s, several researchers (Balota, 1983; Fowler, Wolford, Slade, & Tassinary, 1981; Marcel, 1983) reported that visually obscured words ("masked primes") were cognitively processed outside of conscious awareness. Evidence was offered in the form of these primes' effects on speed and/or accuracy of rapid classification responses to immediately following visible targets. When the masked prime and the immediately following target (word) were semantically related, classification responses to targets were found to be faster and/or more accurate than when there was no semantic relation.

The conclusion that these subliminal priming experiments established an unconscious form of cognitive processing was vigorously contested (Holender, 1986) on grounds that the supporting experiments had not adequately demonstrated that masked primes were processed without conscious awareness. Methods introduced in the 1990s eventually succeeded in demonstrating that unconscious priming could occur not only in the absence of awareness of the masked primes, but even when their conscious visibility was reduced to zero (Greenwald, Klinger, & Schuh, 1995; Greenwald, Draine, & Abrams, 1996; Draine & Greenwald, 1998). Unconscious (or subliminal, or masked) priming subsequently became a widely—even if not universally—accepted empirical phenomenon.

The research reported in this article started when, in 2006, Anthony G. Greenwald suspected that human conditioning could be investigated using methods that had previously been used to investigate subliminal priming. The research proceeded without collaborator until 2010, when Greenwald's and De Houwer's encounter at a conference provided occasion for Greenwald to tell De Houwer about the research already conducted. De Houwer's published skepticism regarding unconscious conditioning (Mitchell, De Houwer, & Lovibond, 2009) prompted Greenwald to invite De Houwer to collaborate. In turn, De Houwer made clear that willingness to be an eventual coauthor would require being able to replicate unconscious conditioning with human subjects in the Ghent University laboratory. That happened in 2013. In 2013 and 2014, further collaborative research on conscious conditioning proceeded, using procedures differing from the unconscious conditioning experiments by extending the duration of the CS—to make it easily visible.

This article reports two series of studies, one on unconscious conditioning (Series 1) and one on conscious conditioning (Series 2). All studies started with a conditioning phase during which a conditioned stimulus (CS) preceded an unconditioned stimulus (US) on each trial. Subjects were asked to categorize the USs (as positive or negative words or as male or female names). The CSs were meaningless letter strings that were followed consistently by US stimuli from one category (e.g., one CS letter-string was always be followed by positive words whereas another CS letter-string was always followed by negative words). A subsequent conditioning test phase presented trials that ended the contingency that characterized the conditioning phase. For instance, a CS letter-string that was always followed by a positive word during conditioning could be followed by a negative word during test. Learning was indexed by the difference in performance on contingency-consistent test trials versus contingency-inconsistent test trials. In the 14 experiments of Series 1, CS letter-strings were presented masked to greatly reduce their visibility. After the test phase, a visibility test was used to confirm effectiveness of this masking. Because visibility of the CS is a prerequisite of awareness of the CS–US contingency, the studies of Series 1 served to investigate the possibility of unconscious conditioning. In the six studies of Series 2, CS letter-strings were easily identifiable, allowing examination of conscious conditioning.

All of the Series 1 studies used the following four elements of method from studies on subliminal priming: (a) a *response window* method that, by obliging very rapid responding, magnified effects of visually masked prime stimuli on responses to immediately subsequent target stimuli, (b) a "sandwich" visual masking procedure that used both forward and backward pattern masks to reduce visibility of prime stimuli, (c) extensive forced-choice testing to document visibility of masked primes, and (d) a statistical regression intercept method (Greenwald et al., 1995; Klauer, Greenwald, & Draine, 1998) that allowed estimation of the magnitude of priming associated with zero visibility of masked primes. Because of the large number of studies reported in this article, presentation of all results is preceded by the General Method section, which describes the main commonalities and differences among the Series 1 experiments. The main difference of method in the Series 2 experiments was lengthening of the CS–US interval, to afford visibility of the CS.

General Method

Subjects and Criteria for Retaining Data for Statistical Analysis

Except for one experiment conducted at Ghent University, subjects in all experiments were undergraduate students at University of Washington whose participation partially fulfilled a requirement of their introductory psychology course. Reported analyses are based on samples that, for all experiments, were reduced by applying criteria (the same in all experiments) to identify subjects who inadequately adhered to task instructions. "Inadequate adherence" included (a) responding too rapidly or too slowly for the response window procedure to be useful,[2] (b) rates of response alternation (or its complement, response repetition) that varied by 2.5 *SD* or more from the expected 50% rates for either condition-

[2] Subjects who cooperate well with the response window instructions have most latencies between 300 ms and 600 ms. It is known that responses with latencies outside this range do not effectively capture effects of masked primes (see Greenwald, Abrams, Naccache, & Dehaene, 2003). Subjects with more than 25% of responses outside this range were dropped. For retained subjects, individual trials outside the range of 250 ms to 800 ms were dropped.

ing or visibility tests, or (c) responding on one of the two response keys at a proportion more than 2.5 *SD* deviant from the expected 50% rate. These nonadherences could have been due to inattention to instructions or to deliberate noncompliance—a strategy some subjects may have used to shorten the experimental session. These criteria eliminated approximately 10% of subjects per experiment. The total numbers of subjects before and after exclusions are shown in Appendix Table A1. The nonadherence criteria were ones used in previous subliminal priming experiments having procedures similar to those used in the present research. Analyses that retained the excluded subjects invariably revealed findings similar to those reported in this article, but statistically weaker because of added dependent measure noise associated with weak adherence to either (or both) the challenging speed instructions of the response window procedure or the perceptually challenging instructions to classify visually masked stimuli that were effectively invisible for most subjects.

Apparatus

Up to five subjects participated concurrently, each in a cubicle with a 42-cm (diagonal dimension) color CRT display operating with 120 Hz refresh rate and a standard *QWERTY* keyboard, both controlled by desktop computers using Inquisit laboratory software (available at http://millisecond.com). Left-hand responses during acquisition, conditioning test, and visibility test phases were made by pressing the *D* key, and right-hand responses were given with the *K* key. Stimuli were presented in Arial bold font with size selected so that uppercase letters occupied 5.3% of screen height. An electric fan motor in subjects' cubicles produced a relatively low level of background white noise to mask extraneous sounds.

Stimuli and Subalphabet Method

The 26-letter English alphabet was divided into two mutually exclusive letter sets (Set A = *ADFHJLNOPWY*; Set B = *BCEGIKMQRSTUVXZ*). Two sets of gendered names and valenced words were constructed for use as USs, one set restricted to letters of each subalphabet. Similarly, two 4-letter consonant string CSs were created from each subalphabet (e.g., *NPLW*, *DHJF* from Set A; *BMVZ*, *GKQX* from Set B). (All stimuli are presented in Appendix Table A2.) The subalphabet method assured that no CS from one subalphabet ever shared a letter with any US (gendered name or valenced word) that had been created using the other subalphabet. The subalphabet strategy thereby avoided complications (in terms of priming) that could result from stimulus overlap between CS and US items. Assignment of a specific subalphabet to CS versus US role was counterbalanced in each experiment except for three experiments (S6, S10, and S15 in Tables A1 and A2) that used just Subalphabet B for CSs and Subalphabet A for USs.[3]

CS consonant strings were always presented as uppercase letters; US words and names were always lowercase. This case difference avoided any possibility of confusion between CS letters and US letters during tests of CS visibility (see the section Visibility Test Trials for Masked CSs, below).

Acquisition Procedure

Subjects' task during acquisition (also conditioning test) trials required speeded categorization of fully visible US words. The USs were either affectively polarized words that were to be categorized as pleasant versus unpleasant in meaning, or they were familiar first names that were to be categorized as male versus female. These stimuli have strongly overlearned valence or gender associations that are widely understood to be activated automatically on perceiving the words or names.

In the acquisition phase, each US word was preceded by one of two CSs, which were 4-letter uppercase consonant strings (e.g., *DHJF* and *GKQX*). The four CS letters were sometimes presented in a single fixed order and sometimes in four randomly permuted orders. (See Table A2 for these details.) The acquisition phase of conditioning consisted of between 192 and 336 CS–US pairings that, in all experiments, were 100% consistent in the pairing of a specific CS with one of the two US categories. That is, one of the 4-consonant CSs (e.g., *DHJF*) always preceded words from one of the US categories (e.g., male names), and a different CS string (e.g., NPLW) always preceded words from the complementary US category (e.g., female names). This is a *differential* conditioning procedure, which provides distinct CSs for each of two complementary USs.

In two of Series 1's experiments, CS duration was manipulated on a within-subjects basis. Experiment 1a provided data for conditioning when a 75-ms CS duration was used in acquisition with either 25 ms or 75 ms CS duration in conditioning test. Experiment 1b provided data for all four combinations of 25-ms and 75-ms CS durations in acquisition and in conditioning test. All other experiments used the same CS duration in acquisition and conditioning test.

All stimuli were presented at screen center (see Figure 1 for a schematic depiction of an acquisition trial). For Series 1's unconscious conditioning experiments each acquisition trial started with a 500-ms medium-gray plus sign ("+") as a focus point. The focus point was replaced by a 300-ms forward mask, which was a randomly selected one of eight patterns of randomly arrayed letter fragments. The forward mask was then replaced by the CS, a consonant string. CS durations ranged between 25ms and 75ms in Series 1. When CSs were briefer than 75ms, a backward masking letter-fragment pattern, randomly selected from 8 mask patterns—with the constraint of being different from the forward mask pattern—followed the CS and completed the CS–US interval. The US word or name started immediately after completion of the 75-ms CS–US interval. In the first block of acquisition trials, the US remained on-screen until the subject responded by pressing the left (*D*) or right (*K*) key to classify the stimulus. After the subject's key-press response, a 600-ms intertrial interval preceded the next trial's starting focus ("+") stimulus.

[3] This subalphabet method has, to date, been developed only for use with gender and valence categories. Those two domains were relatively easy to use because of the large number of available valenced words and gendered names in English language.

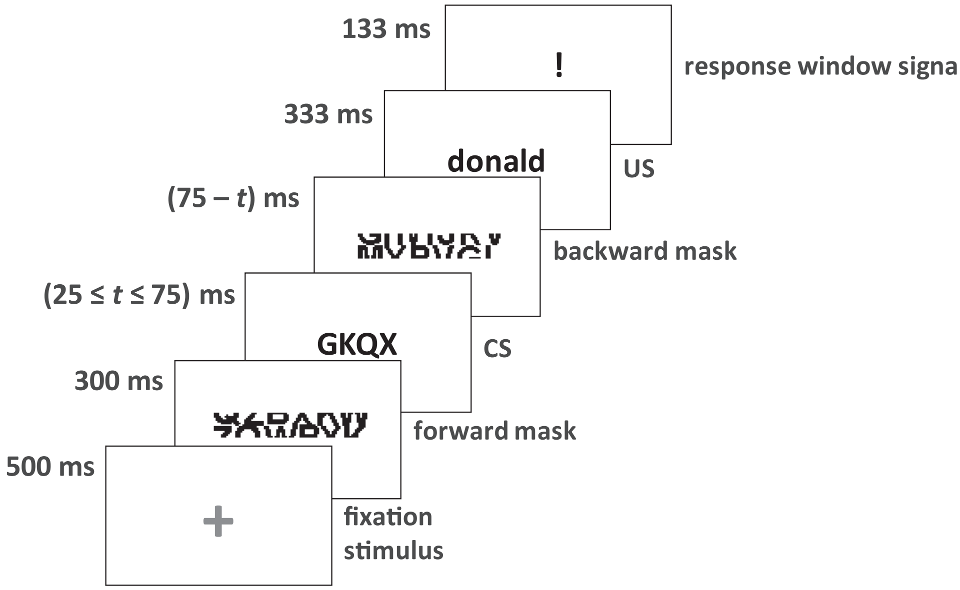

Figure 1. Stimulus presentation sequence in conditioning test trials for Series 1 experiments. All Series 1 experiments used a 75-ms onset interval between starts of CS letter strings and US words. CS duration (t) varied between 25ms and 75ms. For $t < 75$, a backward mask filled the remainder of the CS–US interval. The requirement to respond during the response window signal (exclamation point) obliged sufficiently rapid responding to produce substantial error rates. Conditioning was detected as greater accuracy when the conditioning-test CS–US pairings were consistent with those used during acquisition than when they were inconsistent.

Response Window Procedure

In the second acquisition block and in all subsequent acquisition and conditioning test blocks, a *response window* procedure obliged faster responding to the US than would have occurred with standard ("respond rapidly") reaction time (RT) instructions. The "window" interval, during which the subject was asked to try to respond, was marked by center-screen appearance of a black exclamation point that replaced the US word. Instructions before the second acquisition block stressed the importance of responding *during* the exclamation point's on-screen presence. The exclamation point's duration ("window width") was always 133 ms. Initially (second acquisition block) the exclamation point replaced the valenced word US or gendered name US after that word or name had been on screen for 408 ms. If the subject responded during the exclamation point's presence, the exclamation point turned red and remained red for 300 ms. If the subject failed to respond during the exclamation point's 133-ms on-screen presence, the exclamation point disappeared with no color change. The subject could therefore know that his or her response (a) was faster than desired if the exclamation point never appeared, or (b) was slower than desired if the exclamation point disappeared without turning red. Subjects received end-of-block feedback reporting the percentage of trials for which their response was successfully "in" the window. The requirement to respond rapidly was increased gradually so that the "center" of the 133-ms window advanced from 475 ms after US onset (2nd acquisition block) to 450 ms (typically in the 3rd acquisition block), and to 400 ms (typically in the 4th acquisition block), remaining at that value for subsequent acquisition blocks and for all conditioning test blocks. In this final configuration, the exclamation point started at 333 ms after US onset and ended 467 ms after US onset, or earlier if the response had by then already occurred. This procedure obliged responding about 150 ms faster than most subjects would routinely give a "rapid" response. The resulting error rates were typically above 15%.

Correct responses to the US initiated a standard 600-ms intertrial interval. However, during all acquisition blocks, incorrect responses triggered a 200-ms center-screen appearance of "ERROR" in red capital letters, extending the intertrial interval by that duration. No error feedback was given in subsequent conditioning test trials.

Conditioning Test Trials

Conditioning test blocks were presented with no announcement or appearance of a change from the acquisition procedure, apart from informing subjects that error feedback was being discontinued. Subjects in the Series 1 experiments were typically unaware of the essential procedural change that enabled the conditioning test—instead of being consistently contingent on CS stimuli as in acquisition, US stimuli in conditioning test blocks were presented at random in relation to CSs. Therefore, only 50% of conditioning test trials were consistent with the acquisition contingency; the remaining 50% presented a reversal of the consistent acquisition contingency. A higher error rate on inconsistent than consistent trials would be the indicator that the CS had acquired some influence over the subject's responding.

Visibility Test Trials for Masked CSs

Implementing the advice of Reingold and Merikle (1988), visibility trials used the same stimulus sequence that subjects had experienced during acquisition blocks, including the response window's exclamation point (which no longer turned red, however). The instructed task in visibility test trials was to classify an uppercase consonant string as one of the two 4-letter CS strings that were presented in the trial's sequential position of the CS. Subjects were to use the right (*K*) key to indicate the consonant string that had been associated with right-key responses to US words, and the left (*D*) key for the consonant string that had been associated with left-key responses to US words. Because these consonant strings had not been visible to most subjects previously in the experiment, it was necessary to use two preliminary practice trial blocks to acquaint subjects with the temporal position of the uppercase CS string in the visibility test stimulus sequence—immediately following the forward mask and preceding either the backward mask (if any) or the US (if CS duration was the full 75-ms duration of the CS–US interval). In the two practice blocks, the CS's duration was extended to make it easily visible (133 ms in the first preliminary block and 100 ms in the second). Subsequent visibility test trials were conducted with the CS duration (between 25 ms and 75 ms) and backward mask (if any) that subjects had encountered during their acquisition trials. For CS-visibility test trial blocks, subjects were asked to delay responses until after the response window's exclamation point had disappeared. This instruction was needed because subjects who responded during the exclamation point might have been responding based on a conditioned association rather than based on perception of the CS. To assure that subjects would delay response, any key-press responses that occurred sooner than 333 ms after the exclamation point's disappearance—which was 800 ms after US onset—were ignored, meaning that subjects were obliged to repeat their CS-classification response if their initial response had been made too rapidly.[4]

Contingency Awareness Tests

Even though it was not likely that subjects would be able to learn the contingent CS–US relation in Series 1 experiments, contingency awareness tests were administered in six of the 14 experiments. The contingency awareness test presented four 4-letter consonant strings, each in a separate test item. Two of the four strings had appeared during the subject's acquisition trials and the other two had not been seen at all. (Those previously unseen ones had served as CSs for other subjects in the counterbalancing design.) For each of the four 4-character strings, subjects answered a single question. As illustration, in a condition that used gendered names as USs, one of the four questions was: "In the trials completed so far, did the set of letters *DHJF* usually appear before male or female names?" There were seven response options: "before male names (certainly)," "before male names (probably)," "before male names (guess)," "*DHJF* did not appear in the task," "before female names (guess)," "before female names (probably)," "before female names (certainly)." For a subject who had seen *DHJF* consistently paired with female names, the 7 options were scored ranging from −3 for the diametrically wrong response ("before male names (certainly)") to + 3 for the fully accurate response "before female names (certainly)"). Scores were converted to an awareness score by averaging the accuracy scores for the two strings that the subject had encountered. A fully correct response therefore received a + 3 score. A negative average score indicated guessing in favor of incorrect contingencies.

Signal Detection Sensitivity (d') Measures of Conditioning and CS Visibility

The regression intercept tests described in the next paragraph require that zero values of both CS visibility and conditioning test measures have rational zero interpretations. These rational zero values were assured by converting the data of conditioning and visibility test trials to signal detection sensitivity (d') measures, for which zero scores indicate absence of effect. Use of the d' metric required scoring each trial's response as a 'hit' (vs. a 'miss') or as a 'false alarm' (vs. a 'correct rejection'). When the presented CS was the one associated in acquisition with a right-key response, the hit measure was scored as "1" when subjects gave the (correct) right-key response, and "0" (= miss) otherwise. When the presented CS was one that had been consistently associated in acquisition with the US corresponding to the left key response, the false alarm measure was scored "1" when subjects gave an erroneous right-key response, and "0" (= correct rejection) otherwise. Hit (*h*) and false alarm (*fa*) rates were converted to signal detection d' values by converting their representation as proportions (i.e., hits divided by number of right-key CS trials, and false alarms divided by number of left-key CS trials) to *z* values via an inverse normal transformation, then computing $d' = z_h - z_{fa}$.[5] Because *h* and *fa* proportions of either 0 or 1 do not allow the inverse normal transformation (i.e., they would require a division by zero), an end-point correction was used: 0s were replaced with .25/N and 1s were replaced with (1 – .25/N), where *N* is the number of trials used in the denominator of the proportion. Banaji and Greenwald (1995) found that these end-point conversions were superior to those more standardly recommended in statistics texts (e.g., Agresti, 1990).

For the visibility test, hit rates were computed as proportions of (correct) identifications of the CS associated during acquisition with correct right key responses to USs, and false alarms as proportion of (erroneous) identifications of the CS associated with correct left key responses as the right-key CS.

Intercept Tests in Regressions of Conditioning on CS Visibility and Contingency Awareness

When both the indirect measure of a masked stimulus's effect (conditioning in these experiments) and a direct measure of the

[4] Previous subliminal priming research using the response window procedure had shown that responses slower than 500 ms showed no evidence of influence by a preceding masked prime (Greenwald et al., 2003). This was the reason for wanting to assure that CS classification responses on visibility test trials would be delayed until well after 500 ms had elapsed following appearance of the US name or word on visibility test trials.

[5] To translate conditioning test d' values to more meaningful numbers, d'' values of 0.10, 0.15, and 0.20 can be understood, respectively, as 4%, 6%, and 8% greater accuracy on conditioning-consistent test trials, compared with inconsistent ones. When these d'' values are reported as intercept effects, that differential accuracy measures the level of conditioning associated (in the regression) with performance at chance on the visibility test (i.e., visibility $d'' = 0$).

masked stimulus's visibility are measured on scales with rational zero values, the regression of the indirect effect on the visibility measure estimates the magnitude of indirect effect associated with the value of 0.0 on the direct measure of visibility (Greenwald et al., 1995, 1996). In the present studies, visibility and contingency awareness tests provided direct measures of CS visibility and contingency awareness, respectively. The conditioning d' measure assessed an indirect (uninstructed) effect of the CS. Significantly positive values of the d' measure of conditioning associated with zero on the CS-visibility measure (i.e., the regression intercept) therefore indicated unconscious conditioning—that is, finding a significant conditioning effect associated with the zero value of CS visibility (Greenwald et al., 1995). Similarly, significantly positive conditioning values associated with zero on the contingency awareness measure would indicate a different form of unconscious conditioning—conditioning in the absence of awareness of the CS–US contingency. In this research, positive intercept d' values for which 99% confidence intervals excluded zero were interpreted as indicating conditioning in the absence of visibility or awareness. A 95% confidence interval was not used because the number of such tests conducted in this research (e.g., 18 tests in the Series 1 experiments) afforded excessive probability of Type I error due to the use of multiple tests—a Bonferroni correction procedure might have justified using an even wider confidence interval than 99%.

An important caveat about these regression intercept tests is that, when the predictor (either visibility or contingency awareness) has imperfect reliability the intercept can be spuriously statistically significant. This possibility was considered in detail in this research, aided by the simulations described under the next heading.

In the two Series 1 experiments in which CS duration was manipulated, separate regression analyses were conducted for each combination of acquisition and conditioning test CS durations. Visibility d' measures for these regressions were computed using the subset of visibility test trials based on just the visibility test CS stimuli that had the same experimental history (acquisition CS duration and conditioning test CS duration). These CSs were presented in visibility tests at the same CS duration used for them in the conditioning test.

Simulations to Generate and Test Spuriously Significant Intercept Effects

It is not difficult to generate data sets that have spuriously significant intercepts. Simulations designed with (a) true criterion-measure intercepts of zero, (b) a positive slope of the criterion measure on the predictor, and (c) a predictor with nontrivial measurement error (unreliability) will routinely produce spuriously significant intercepts if the regression slope is more than weakly positive. A simulation with such characteristics was used to produce spuriously significant intercepts in 10 simulated samples of $N = 110$ (this was the average sample size for the nine Series 1 experiments that used CS durations ≥58 ms). The statistical program to generate these samples, along with an Excel spreadsheet that rapidly produces and graphically displays simulations with spurious intercepts can be found in the study archive (to be made publicly available together with publication of this article). An unusual, but desirable, feature of these simulations is that they allowed measurement errors for both predictor and criterion variables to vary as a function of the latent ("true-score") values of these variables. To explain why this was done: (a) random error necessarily comprises 100% of a conditioning or visibility measure when the latent value for that measure is zero and (b) random error logically should decrease to zero as scores of each of these measures achieve their maximum values (indicating perfect visibility or contingency awareness). Intuitively, this simulation feature maps onto an assumption that (a) when true visibility or true contingency awareness is totally absent, observed visibility or awareness scores are purely error variance, and (b) this error component should decrease to zero as these latent measures achieve maximum values. The first 10 data sets produced by the simulation script that incorporated these features were saved for analyses that are described in this article; these analyses used the errors-in-variables adjustment method described by Klauer, Draine, and Greenwald (1998) and Klauer, Greenwald, and Draine (1998).

Power and Data Collection Stopping Strategies

An aim of this research was to establish limiting conditions of unconscious conditioning findings. The initial findings of Series 1 experiments suggested that CS duration was a critical moderator, which explains why the Series 1 experiments examined a range of CS durations. The most important inferential statistical tests were for significance of a regression intercept effect that, if significant, supports a conclusion of unconscious conditioning. Subliminal priming experiments that had used the same form of regression intercept test usually observed intercepts of $d' = 0.10$ (or larger), with standard deviations of intercept estimates typically close to 0.25. Using those as the basis for power calculations, experiments with $N = 70$ had power of .92 (at 2-tailed $\alpha = .05$) to find a statistically significant finding if a true intercept value was 0.10. Most experiments were therefore conducted with at least 70 subjects. The one experiment conducted at Ghent University had 32 subjects. For practical considerations, this was an initially planned stopping point. Using the observed effect size ($d' = 0.177$) of the experiment for which it was a replication, and using an expected standard deviation of 0.30, the power associated with sample size of 32 was .92.

Power considerations were less central for experiments in which the aim was to establish limiting conditions for the conditioning effect (i.e., when null results were expectable). To establish a null conclusion, the need is to identify a narrow confidence interval that includes the intercept value of zero. This required considerably larger sample sizes, which were used especially for the present Series 2 (conscious conditioning) experiments.

In most of the present experiments, data collection was stopped at a point too early for appropriate hypothesis testing—this was a standard laboratory procedure to assure that experiments were free of previously undetected programming errors. In a few cases in which program errors were indeed found, flawed data were discarded and the experiment was reinitiated with corrected procedures.

For several experiments that ultimately had null findings, data collections were reinitiated after an initially planned stop. The reinitiation was in the expectation of allowing increased confidence in a null finding. In one experiment for which analysis after the initially planned stop yielded a statistically significant result with a p value falling in an understood 'marginal' range, additional data were collected to increase confidence that the result warranted interpretation as a null hypothesis rejection. A few experiments were run with more subjects than needed for adequate power, for the (scientifically irrelevant) reason that it was administratively inconvenient to stop data collection.

Series 1 (Unconscious Conditioning) results. Series 1 provided 18 hypothesis within 14 experiments. Seven of the 18 hypothesis tests are presented here as Experiments 1a (2 hypothesis tests from within-subject conditions), 1b (4 hypothesis tests from within-subject conditions), and 1c (1 condition). Results for 11 additional Series 1 hypothesis tests are in the online supplemental material. These additional experiments served partly to replicate results in Experiments 1a, 1b, and 1c, and partly to locate a boundary between acquisition CS durations (≤42ms) too brief to obtain unconscious conditioning and durations (≥58 ms) at which conditioning was reliably observed.

Experiment 1a used a 75-ms duration for CSs on all acquisition trials, while varying CS duration randomly between 25 ms and 75 ms, within blocks, on conditioning test trials. The data, which are presented in Figure 2, show that conditioning was observed only when the 75-ms CS duration was used in the conditioning test (Panel A of Figure 2). The same subjects, tested with the same stimuli, but presented in conditioning test with a 25 ms CS duration (plus 50 ms backward mask) showed no evidence for conditioning (Panel B of Figure 2).

Experiment 1b had four within-subject conditions, with CS duration varied between the two values of 25 ms and 75 ms during *both* acquisition and conditioning test trials— randomly within blocks of trials in both phases. This procedure allowed CSs experienced only at 25-ms duration during acquisition trials to be tested during conditioning tests at either 25-ms or 75-ms durations. Similarly, CSs experienced with 75-ms duration in acquisition trials were tested at both 25-ms and 75-ms durations in the conditioning test phase. This design required use of distinct CS stimuli at the two durations during acquisition (the stimuli are shown in Table A2). Experiment 1b's findings (see Figure 3) showed that conditioning occurred only when the longer CS duration (75 ms) was used for *both* acquisition and conditioning test phases. That is, the only one of the four within-subject treatments that produced conditioning involved stimuli experienced with 75 ms CS duration in acquisition and then tested with 75 ms CS duration during the conditioning test phase.

Additional experiments in Series 1 (see Figure 4) had CS durations of 25, 33, 42, 58, or 75 ms, used identically in both acquisition and conditioning tests. (Procedural details for all of these are in Appendix Tables A1 and A2.) One other replication examined specifically the combination of 75 ms CS duration in acquisition and 25-ms duration in conditioning test, which appears in both Experiment 1a (Figure 2, Panel B) and Experiment 1b (Figure 3, Panel C). Because the result for this combination in Experiment 1b was marginally significant, the added replication (see online supplemental material, Figure S2) was useful to reinforce the null conclusion for this combination of CS durations in acquisition and conditioning test phases. Figure 4's summary of the Series 1 findings makes apparent that conditioning was evident only in the nine experiments for which CS durations were 58 ms or greater in *both* acquisition and conditioning test. In only one of those nine experiments was the key statistical test (for the intercept effect) nonsignificant at the .01 level (see online supplemental material, Figure S11).

Experiment 1c (see Figure 5) was conducted at the end of Series 1, using 75-ms CS durations in both acquisition and conditioning test phases. This combination of CS and US durations had already

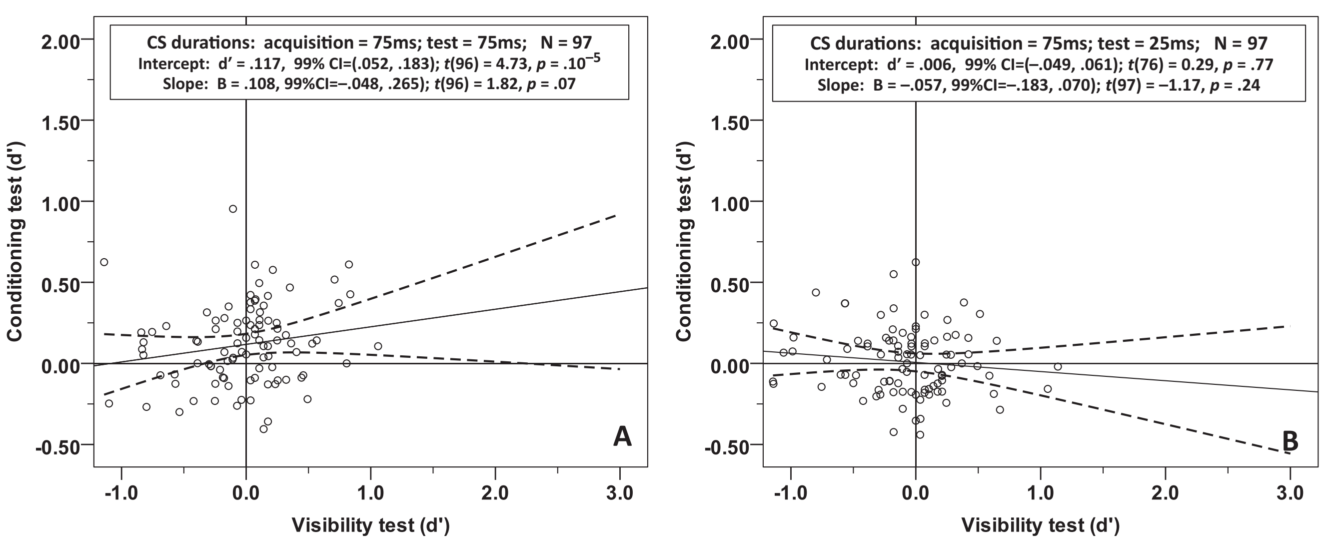

Figure 2. Results of Experiment 1a. (A) When the CS–US interval in both acquisition and conditioning test was 75 ms, unconscious conditioning was obtained, indicated by positive difference from zero of the intercept in the regression of the conditioning measure on the CS-visibility measure. (B) No conditioning was apparent when the CS–US interval in conditioning test was reduced to 25 ms. Each data point (small circle) indicates a single subject's data. Dashed lines bound the regression slopes' 99% confidence intervals. Experiment 1a had a within-subjects design in which CS durations were varied within conditioning test blocks. As a consequence, each subject (N = 97) appears in both panels.

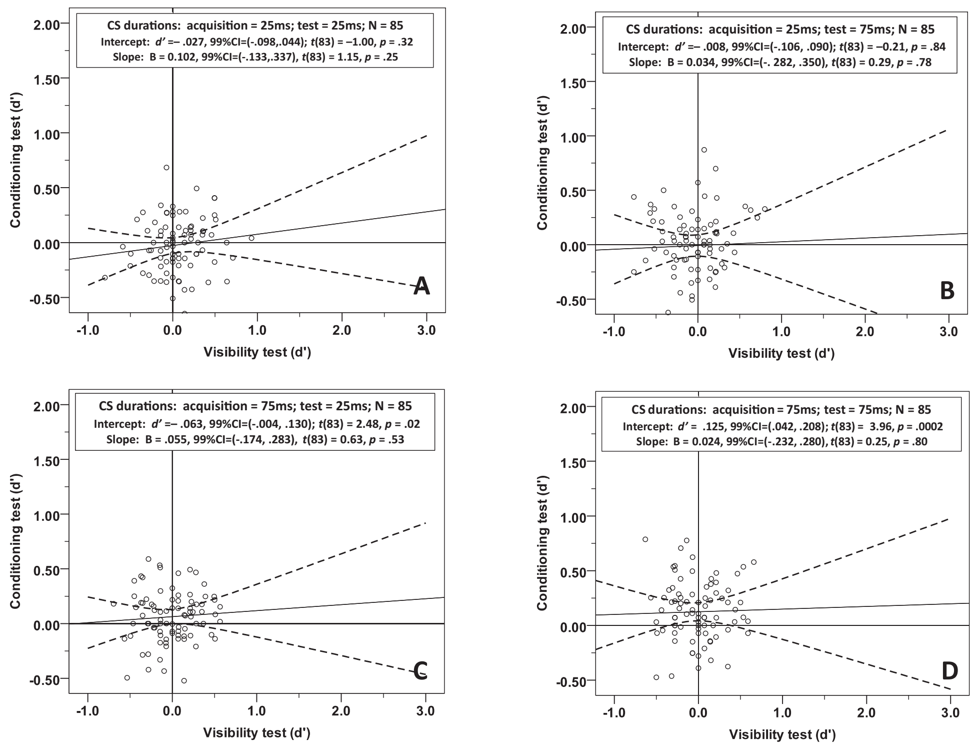

Figure 3. Results of Experiment 1b. Interpretation of regression intercepts is as in Figure 2. Only in Panel D, where CS duration in both acquisition and conditioning test trials was 75ms (rather than 25ms in either or both) was there evidence for unconscious conditioning. Dashed lines bound the regression slopes' 99% confidence intervals. Experiment 1b had a within-subjects design in which CS durations were varied within-blocks. As a consequence, each subject ($N = 85$) appears in all four panels.

been shown to produce conditioning in Experiments 1a and 1b, as well as in three other replications (see the online supplemental material, Figures S15–S17). Experiment 1c was the largest of the six Series 1 experiments that included a contingency awareness measure. The regression scatterplot in Figure 5 shows that scores for contingency awareness in Experiment 1c were symmetrically distributed around zero. Not even one subject in Experiment 1c showed perfect contingency awareness. Nevertheless, conditioning was statistically significant. Experiment 1c's test for regression of conditioning performance on CS visibility (rather than contingency awareness) shows the usual Series 1 pattern of a significant intercept effect and a nonsignificant regression slope (see scatterplot in the online supplemental material, Figure S18).

Series 2 (Conscious Conditioning) Method and Results

The two experiments in Series 2 allowed CSs to be visible to all subjects, by using a 375ms CS–US interval and omitting the forward mask used in all Series 1 experiments. CSs, which ranged in duration from 75 ms to 375 ms in these experiments, started immediately after the trial-initiating 500-ms focus ("+") stimulus (no forward mask for the CS was used).

Each of Experiments 2a and 2b consisted of three subexperiments that varied CS presentation procedures to vary ease of noticing and remembering CS–US contingences. Experiment 2a had more acquisition trials (250) than Experiment 2b (168). Experiments 2a and 2b also differed in the location of their contingency awareness tests in the subjects' series of tasks (see General Methods section for details). Measuring contingency awareness after the conditioning test in Experiment 2a avoided possible influence on the conditioning test of contingency information that might be learned during the awareness test. Placing the contingency awareness test prior to the conditioning in Experiment 2b maximized likelihood of measuring contingency awareness before possible forgetting. Use of both sequences provides more confidence in findings than if only one of these had been used.

The differences among three variations of each experiment assured that there would be a range of contingency awareness scores that extended to the extremes of the measure—ranging from zero contingency awareness to perfect contingency awareness. This strategy was needed to obtain adequately powerful regression tests of the role of contingency awareness in conditioning.

In Experiment 2a, greatest awareness was produced by giving advance information of the specifics of the contingency (i.e., the details of which 4-letter CS consonant string was associated with which US word or name category); intermediate awareness used the same stimulus presentation procedure without providing any advance contingency information; and lowest contingency awareness was produced by additionally reducing the CS from 375-ms duration to the first 83 ms of that interval, followed by a backward mask for remaining 292 ms of the CS–US interval. Because of the

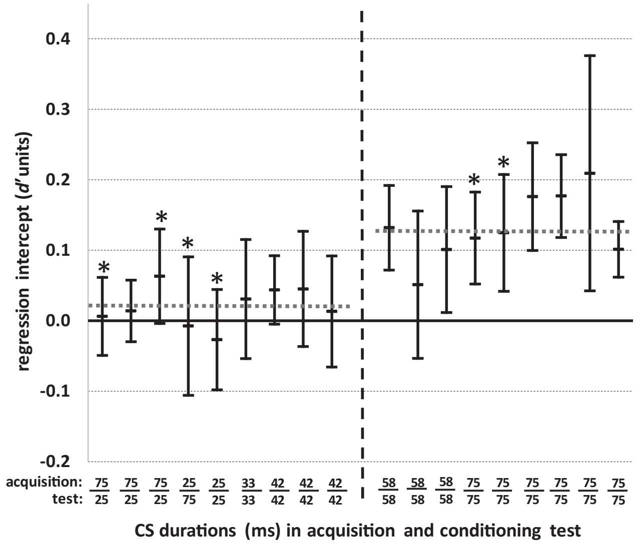

Figure 4. Summary intercepts for regressions of conditioning on CS visibility for the 18 data sets in Series 1. Error bars are 99% confidence intervals. The dashed vertical line separates 9 data sets for which CS durations were ≤42 ms from 9 data sets for which CS durations were ≥58 ms. Dotted horizontal lines show weighted means of the intercept for each of these groups of data sets. Asterisks mark the six data sets for which individual subject data are displayed in Figures 2 and 3. The 18 data sets are in the same order (left to right) as the 18 regression scatterplots in the online supplemental material (Figures S1–S18).

lack of any forward mask, the CS was clearly visible to all subjects even with the backward-masked 83-ms CS.

Because the three variations of Experiment 2a, overall, observed greater contingency awareness than desired (too many subjects with perfect contingency awareness), the three variations in Experiment 2b mildly increased the difficulty of contingency discovery. To obtain the highest level of contingency awareness in Experiment 2b, the CS was presented for full duration of the 375-ms CS–US interval in acquisition and subjects were given advance information that there was a contingency in effect, but they were not given the specifics of the contingency; for the intermediate level, subjects received no advance information about presence of a contingency and the CS duration was reduced to 75 ms followed by a blank screen for 300 ms (the CS remained fully and easily visible); to produce the lowest level of contingency awareness, the 75-ms CS duration was followed by a 100-ms backward mask, which was followed by a blank screen for 200 ms. With this last procedure, the CS remained fully visible, but it had no visible persistence, creating an increased processing burden to retain awareness of the CS's identity until the US was presented. Of the 583 subjects who completed awareness tests in Experiments 2a and 2b, 37% had zero contingency awareness, 31% had perfect contingency awareness, and the remaining 32% were distributed between these extremes.

The procedure variations that reduced contingency awareness in Experiments 2a and 2b did not interfere with clear visibility of the CS. Those who became aware of the contingency during acquisition were expected to discover rapidly during the conditioning test phase that the contingency was no longer in effect. Unsurprisingly, therefore, evidence for conditioning weakened from the first to the second of the two 48-trial conditioning test blocks. Figure 6 shows that there was a significant positive relationship between contingency awareness and conditioning that extended to the second conditioning test block in Experiment 2a. Figure 7 shows that in Experiment 2b, which overall had less contingency awareness than Experiment 2a, the positive relation between contingency awareness and conditioning was limited to the first conditioning test block.

The finding that all four regression intercepts in Figures 6 and 7 were slightly positive (but not significantly so at $\alpha = .05$, 2-tailed) suggested that subjects who lacked contingency awareness might have displayed some weak conditioning effect. However, these weak positive intercepts could also have been statistical artifacts resulting from measurement error in the regression predictor. This artifact possibility can be evaluated with an *errors in variables* method (Klauer, Draine, & Greenwald, 1998; Klauer, Greenwald, & Draine, 1998) that adjusts estimates of both regression slope and intercept for measurement error in the predictor. Before this adjustment, the four intercepts were 0.158, 0.088, 0.053, and 0.045). After adjustment, none of these intercepts was above zero—they were, respectively, −0.476, −0.343, −0.019, and −0.012.

To determine whether conditioning was obtained among subjects whose contingency awareness scores revealed no awareness, the data for subjects having zero values on the contingency awareness measures in the six procedural variations (three in each of Experiments 2a and 2b) were meta-analyzed to provide an overall Bayes Factor estimate of odds favoring the null hypothesis (i.e., no conditioning) relative to the alternative

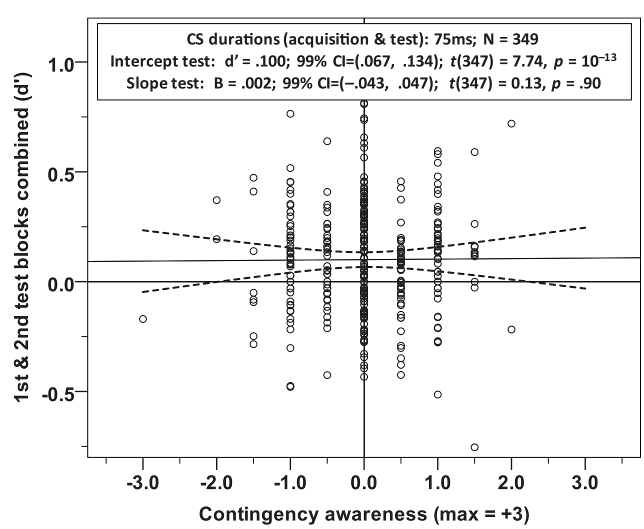

Figure 5. Conditioning test data of Experiment 1c, plotted as a function of CS–US contingency awareness. With the CS duration of 75ms in this experiment, not even one subject achieved a perfect contingency awareness score of + 3. All scores reflect some degree of guessing, which was inaccurate (below zero) as often as accurate (above zero). Dashed lines bound the regression slope's 99% confidence interval.

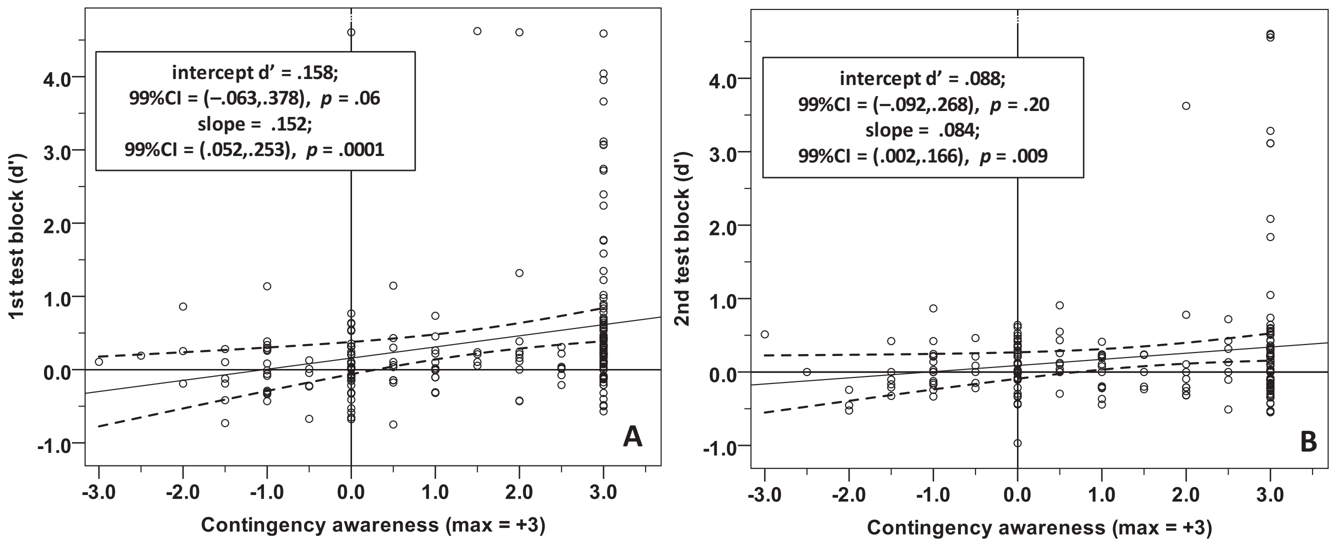

Figure 6. Conditioning as a function of contingency awareness in first (A) and second (B) conditioning test blocks of Experiment 2a. CS–US intervals of 375ms allowed all CSs to be visible. For the awareness measure, +3 indicates perfect contingency knowledge. Dashed lines bound the regression slopes' 99% confidence intervals. The combination of significant slopes and nonsignificant intercepts support the conclusion that conditioning required awareness of the CS–US contingency. Each circle represents a single subject's data ($N = 213$).

hypothesis of conditioning >0.[6] This test was conducted separately on results for the first conditioning test block alone, the second block alone, and the average of both blocks. Results favored the null hypothesis by odds ratios of 21.0:1, 10.5:1, and 19.3:1, respectively, in these three tests. All three odds ratios exceeded the 10:1 criterion for characterizing a Bayes Factor estimate as showing "strong evidence for H_0" (Wetzels et al., 2011).

General Discussion

Series 1 found statistically significant evidence for conditioning in 8 of 9 data sets for which CS durations were at least 58 ms in *both* acquisition and conditioning test phases. In contrast, there was *no* evidence for conditioning in any of nine experiments in which CS durations were briefer than 42 ms in *either* acquisition or conditioning test. Experiment 1c found evidence for conditioning in the absence of evidence either for conscious perception of the CS or contingency awareness. All of the present experiments used a conservative 2-tailed $\alpha = .01$ significance criterion, due to the number of significance tests that were conducted. The left half of Figure 4 shows that two of the nine experiments with CS durations $\leq$42 ms (3rd and 7th from the left) had regression intercepts at the margin of statistical significance. Also, the meta-analytic average of those nine intercepts (dotted horizontal line in the left half of Figure 4) was slightly positive. This is consistent with the hypothesis of a very small conditioning effect—one that would require an extremely large sample size to produce statistical significance; this could be an effect that occurred in only a small minority of subjects.

In Series 2, using fully visible CSs presented during a 375 ms CS–US interval, conditioning effects reliably occurred for subjects who could partially or completely report the CS–US contingencies; however, conditioning was absent for subjects who lacked contingency awareness.

Together, the findings of the Series 1 and Series 2 experiments identify two types of conditioning. One type requires neither conscious perception of the CS nor awareness of the CS–US contingency, and is reasonably labeled 'unconscious conditioning.' The other requires both a visible CS and awareness of the CS–US contingency, and is reasonably labeled 'conscious conditioning.' This identification of two types of conditioning opposes single-process propositional theories of human conditioning, according to which all human conditioning requires awareness of the CS–US contingency. At the same time, it also shows that when CS–US intervals and CS durations are long enough to permit visibility of the CS, conditioning does depend on contingency awareness.

How Convincingly Has Unconscious Conditioning Been Demonstrated?

Regression analyses of the Series 1 experiments with CS durations $\geq$58 ms repeatedly found statistically significant evidence for conditioning when visual discriminability of CSs was at chance. Because the regression intercepts that tested these effects can be inflated when there is measurement error in the visibility predictor (Klauer, Greenwald, & Draine, 1998), those skeptical about unconscious conditioning may suggest the alternative hypothesis that the significant intercepts in the Series

[6] The Bayes Factor analysis used the meta.ttestBF command of BayesFactor 0.9.12–2 (downloaded November 28, 2015, from http://ftp.ussg.iu.edu/CRAN/bin/windows/contrib/3.2/BayesFactor_0.9.12-2.zip).

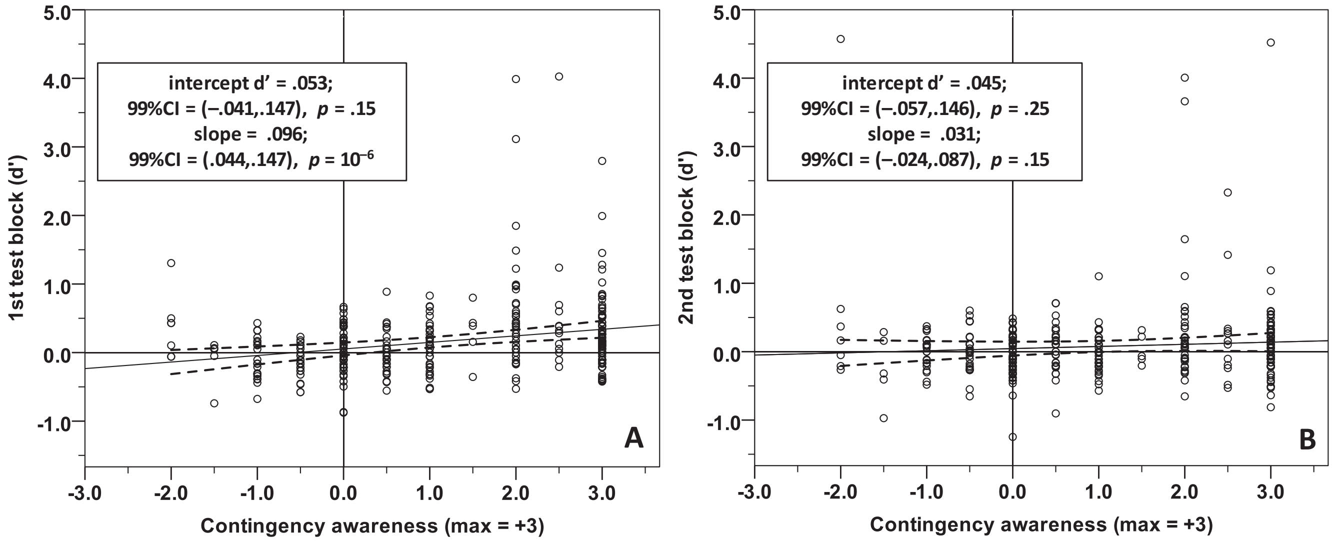

Figure 7. Conditioning as a function of contingency awareness in first (A) and second (B) conditioning test blocks of Experiment 2b. The CS–US interval was 375ms for both acquisition and conditioning test trials. CSs were visible, but were reduced in duration in the second and third of three procedural variations, all of which are combined in this plot. The absence of significant intercepts and the presence of a significant positive slope (A only) is consistent with the conclusion that conditioning requires awareness of the CS–US contingency. Dashed lines bound the regression slopes' 99% confidence intervals. Each circle represents a single subject's data (N = 337).

1 experiments are such artifacts. In more detail, this skeptical argument has four steps: (i) Conditioning requires conscious perception of the CS, which is either sufficient to produce conditioning, or may be needed to allow development of contingency awareness that is needed for conditioning; (ii) the visibility measure contains measurement error and may therefore fail to reveal subjects' true perception of CSs; (iii) therefore, some proportion of subjects who have been conditioned because the CS was consciously visible to them have CS-visibility scores that are (erroneously) near zero; and (iv) these (presumably) consciously conditioned subjects are responsible for the significant intercept findings. Validity of this 4-step argument can be evaluated using three types of data that were available in Series 1's experiments.

Tests of contingency awareness in Series 1 experiments. Experiment 1c was the largest of six Series 1 experiments that included contingency awareness measures. Because of its brief (75 ms), masked CS, subjects in this experiment were not expected to develop contingency awareness. The contingency awareness test was administered immediately after 5 blocks of acquisition trials (total of 192 trials) that had a 100% consistent CS–US contingency. The regression scatterplot of Experiment 1c's conditioning result (see the online supplemental material, Figure S18) shows a statistically significant intercept. Figure 5 shows Experiment 1c's conditioning test measure regressed on its contingency awareness measure. In addition to showing a significant intercept that reveals conditioning in the absence of contingency awareness, Figure 5 shows that not even one subject in the experiment achieved complete contingency awareness. Of the three other experiments in Series 1 that had CS durations ≥58 ms and used the contingency awareness measure, all found significant intercept effects (see the online supplemental material, Figures 10, 11, and 12) and all also showed that not a single subject had complete contingency awareness.

In one of the four Series 1 experiments with CS duration ≥ than 58 ms that had a contingency awareness test (its conditioning data are in the online supplemental material, Figure S10), the contingency awareness test had a different purpose. After three (of five) acquisition blocks, subjects in two of three between-subjects conditions received instructions informing them (a) that there was a contingency in effect, (b) that they should attempt to learn it, (c) that to make it possible for them to learn the contingency, the 4-consonant CSs would be displayed for a sufficient duration for them to be easily visible, and (d) that after 20 such pairings they would be tested for knowledge of the contingency. After those 20 trials (with 375-ms CS durations) and the contingency awareness test, the two remaining blocks of acquisition returned to the previous 58-ms CS durations, after which subjects then continued with the conditioning test followed by the CS visibility test. In one of these two conditions, the instructed contingency (after Block 3 of acquisition) was the one used in all other acquisition trials. In the second condition, the instructed contingency was the reverse (right—left switched) of the contingency used for all other acquisition trials. In each of these two conditions, almost all subjects showed that they had properly learned the contingency presented during the 20 visible-CS trials. The subsequent conditioning test showed that the (properly) learned contingency had no effect at all; it produced neither greater conditioning when it was consistent with all other acquisition trials, nor did it reduce conditioning when it was reversed. Hence, the conditioning effect observed in this study did not depend on what participants consciously learned about the CS–US contingencies. Because of that lack of effect, the results shown in the online supplemental material's Figure S10 are collapsed over the experiment's three conditions.

The small subset of subjects who could see masked CSs showed no enhanced conditioning. Regression plots presented in the online supplemental material include five Series 1 experiments in which CS durations ≥58ms allowed a small proportion of subjects to discriminate the CSs on many or most trials (see data in the online supplemental material, Figures S10 and S15–18). Visibility test d' values of 2.0, 2.5, and 3.0 in those experiments translate, approximately, to correct response rates of 84%, 89%, and 93%. Even higher visibility scores than $d' = 3.0$ occurred occasionally in these experiments, indicating that some subjects could see the CS on nearly every trial. If slight visibility is sufficient for conditioning, it is difficult to understand why the subjects who had those high levels of visibility did not show noticeably stronger conditioning than the much larger group of subjects whose visibility scores were near zero. The appropriate interpretation may be that, even for subjects who could see the CS in Series 1's experiments, the 75-ms CS–US interval was too brief to permit them to process the CS–US relation sufficiently to consciously discern the contingency.

Statistical adjustments supported by simulations indicated that regression intercept effects were not spurious. The statistical procedure that was used in Series 2's experiments to adjust regression intercepts for unreliability of the contingency awareness predictor (Klauer, Draine, & Greenwald, 1998; Klauer, Greenwald, & Draine, 1998) was applied similarly to the nine Series 1 experiments in which CS durations were ≥58 ms. The adjustment algorithm halted for two of these experiments—ones that had visibility scores averaging very near zero; this is understood as a circumstance in which adjustment is not needed. Each of the other seven significant intercept effects remained statistically significant after adjustment; their Cohen's d values averaged 0.34 both before and after applying the adjustment algorithm.

To be confident in the adjustment algorithm's effectiveness, 10 simulation data sets based on true intercepts of zero were generated (see General Methods for details of these simulations). The 10 simulated intercepts ranged, in Cohen's d units, from 0.06 to 0.36. Eight of these were statistically significant. Because of the known model used to generate the data, their statistical significance could only have been spurious. When the Klauer et al. adjustment algorithm was applied, none of the eight that had been (spuriously) significant remained statistically significant (all $p > .18$, 2-tailed). The adjustment dropped the average effect size for the 10 simulated intercepts from Cohen's $d = 0.26$ before correction to Cohen's $d = 0.02$ after correction.[7]

The several data-based considerations make it extremely implausible that the eight statistically significant intercept effects observed in Series 1's experiments were artifacts of unreliably measured CS visibility. Both individually and in the aggregate, those eight significant intercept effects warrant interpretation as indicators of unconscious conditioning.

What Has Been Conditioned to What?

The effective CSs in this research might have been each consonant string as a unit, or the individual letters in each string, or even the component features of those letters. The research was not designed to distinguish those possibilities. Related to this unanswered question is an answerable one that was raised by a reviewer of this article: Could there have been confounds in the design such that letters of the US words might have been more similar to the letters of their paired CS than were letters of the complementary US word category? Because CS and US words were never selected from the same subalphabet (see description of the subalphabet method in the General Methods section) there was no possibility that the letters used in CS strings could be used in any US words with which they were paired in any experiment. And, because assignments of specific CS consonant strings to US categories were counterbalanced in most experiments (see Table A2), those counterbalanced experiments did not allow the possibility of a priming artifact due to shared letter features between CSs and USs from different subalphabets.

Theoretically, the more interesting part of the question about what was conditioned concerns identification of the conditioned response. During conditioning tests, did the CS presentation (a) elicit a conditioned manual response (to left or right key), (b) influence the identification of the specific stimuli used as USs, or (c) influence the identification of the semantic category to which the USs belonged? Future research may be able to provide a decisive empirical answer to this question.

Summary and Theoretical Conclusions

Previous findings of unconscious conditioning in human subjects (e.g., Clark et al., 2002; Clark & Squire, 1998) have often been greeted skeptically. This article's Series 1 experiments avoided objections posed in past critiques by adapting methods that had previously served to overcome critiques of research demonstrating subliminal (i.e., unconscious) priming. Using those methods, this article's Series 1's findings of experiments that used CS–US intervals of 75 ms and CS durations ≥58 ms in both acquisition and conditioning test repeatedly found (in 8 of 9 experiments) statistically significant evidence for conditioning. This conditioning occurred in the absence of evidence for either visibility of the CSs (all 9 studies) or awareness of the CS–US contingency (in the four experiments that tested for contingency awareness).

This article's Series 2 experiments differed from Series 1 in using a longer (375 ms) CS–US interval in which CSs were fully visible. In one of these, contingency awareness tests were placed after conditioning tests, to avoid any possibility that CS information presented in testing contingency awareness might produce contingency awareness that had not been achieved during acquisition trials.

Series 1's unconscious conditioning findings fit with an associative *binding* interpretation—this binding can be understood as

[7] A concern sometimes raised regarding the possibility of spurious intercepts is that a curvilinear function passing through the origin (i.e., no significant intercept) and rising rapidly at low levels of visibility before leveling off might be statistically mistaken for a significant intercept. This hypothesis is not addressed by the Klauer correction method, nor has any additional method been developed for this purpose. No better suggestion is available than to examine regression scatterplots (see the online supplemental material, Figures S10–S18) to see if some such curvilinear function seems consistent with the displayed data.

the integration of features that belong to a single ‘event file’ (Hommel, 2004). Because the likelihood of two features (e.g., CS and US) being integrated into the same neural record should decrease as their temporal separation increases, unconscious conditioning may require very brief interstimulus intervals, such as the 75-ms CS–US interval used in Series 1. This interpretation bears on the hypothesis that unconscious conditioning is more likely with delay than with trace conditioning procedures (Clark & Squire, 1998; Weidemann, Best, Lee, & Lovibond, 2013). In delay conditioning, CS offset is simultaneous with US offset; in trace procedures CS offset precedes US onset. The present findings suggest that brevity of the interval between CS and US onsets is more critical to unconscious conditioning than is simultaneity of their offsets, which did not characterize any of the present experiments.[8]

Even without a decisive mechanistic understanding of differences between conscious and unconscious conditioning, the present findings challenge the assumption that a single-process propositional account can explain all human conditioning. The only substantial change in procedure between the present Series 1 and Series 2 experiments was an increase of the interval between CS and US onsets from 75 ms to 375 ms, thereby rendering the CS visible, which in turn provided opportunity to learn the CS–US contingency. By establishing that unconscious and conscious conditioning can be produced and distinguished using procedures that differ only in the CS–US temporal relation, the present findings provide a basis for further investigations to develop understanding of these distinct forms of conditioning.

[8] To be clear, simultaneous offset of CS and US was not possible with the present procedure, in which the CS and US stimuli appeared at the same (centered) screen location. For offsets to be simultaneous, the two would have had to be superimposed during US presentation, which would have rendered the US illegible.

References

Agresti, A. (1990). *Categorical data analysis*. New York, NY: Wiley.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22,* 88–104. http://dx.doi.org/10.1016/S0022-5371(83)80008-5

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68,* 181–198. http://dx.doi.org/10.1037/0022-3514.68.2.181

Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences, 6,* 524–531. http://dx.doi.org/10.1016/S1364-6613(02)02041-7

Clark, R. E., & Squire, L. R. (1998, April 3). Classical conditioning and brain systems: The role of awareness. *Science, 280,* 77–81. http://dx.doi.org/10.1126/science.280.5360.77

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127,* 286–303. http://dx.doi.org/10.1037/0096-3445.127.3.286

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,* 541–555. http://dx.doi.org/10.1037/0096-3445.113.4.541

Dulany, D. E., Jr. (1961). Hypotheses and habits in verbal “operant conditioning.” *Journal of Abnormal and Social Psychology, 63,* 251–263. http://dx.doi.org/10.1037/h0047703

Fowler, C. A., Wolford, G., Slade, R., & Tassinary, L. (1981). Lexical access with and without awareness. *Journal of Experimental Psychology: General, 110,* 341–362. http://dx.doi.org/10.1037/0096-3445.110.3.341

Greenwald, A. G., Abrams, R. L., Naccache, L., & Dehaene, S. (2003). Long-term semantic memory versus contextual memory in unconscious number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 235–247. http://dx.doi.org/10.1037/0278-7393.29.2.235

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996, September 20). Three cognitive markers of unconscious semantic activation. *Science, 273,* 1699–1702. http://dx.doi.org/10.1126/science.273.5282.1699

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,* 22–42. http://dx.doi.org/10.1037/0096-3445.124.1.22

Hendrickx, H., De Houwer, J., Baeyens, F., Eelen, P., & Van Avermaet, E. (1997). Hidden covariation detection might be very hidden indeed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 201–220. http://dx.doi.org/10.1037/0278-7393.23.1.201

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9,* 1–23. http://dx.doi.org/10.1017/S0140525X00021269

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8,* 494–500. http://dx.doi.org/10.1016/j.tics.2004.08.007

Klauer, K. C., Draine, S. C., & Greenwald, A. G. (1998). An unbiased errors-in-variables approach to detecting unconscious cognition. *British Journal of Mathematical & Statistical Psychology, 51,* 253–267. http://dx.doi.org/10.1111/j.2044-8317.1998.tb00680.x

Klauer, K. C., Greenwald, A. G., & Draine, S. C. (1998). Correcting for measurement error in detecting unconscious cognition: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General, 127,* 318–319. http://dx.doi.org/10.1037/0096-3445.127.3.318

Krasner, L. (1958). Studies on the conditioning of verbal behavior. *Psychological Bulletin, 55,* 148–170. http://dx.doi.org/10.1037/h0040492

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. http://dx.doi.org/10.1037/0097-7403.28.1.3

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197–237. http://dx.doi.org/10.1016/0010-0285(83)90009-9

Mastropasqua, T., & Turatto, M. (2015). Attention is necessary for subliminal instrumental conditioning. *Scientific Reports, 5,* 12920. http://dx.doi.org/10.1038/srep12920

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183–198. http://dx.doi.org/10.1017/S0140525X09000855

Razran, G. (1955). Conditioning and perception. *Psychological Review, 62,* 83–95. http://dx.doi.org/10.1037/h0046875

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863. http://dx.doi.org/10.1016/S0022-5371(67)80149-X

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56,* 97–115. http://dx.doi.org/10.1080/02724980244000297

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44,* 563–575. http://dx.doi.org/10.3758/BF03207490

Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Sciences, 9,* 329–334. http://dx.doi.org/10.1016/j.tics.2005.05.010

Seitz, A. R., & Watanabe, T. (2008). Is task-irrelevant learning really task-irrelevant? *PLoS ONE, 3,* e3792. http://dx.doi.org/10.1371/journal.pone.0003792

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology, 61,* 273–301. http://dx.doi.org/10.1146/annurev.psych.093008.100519

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367–447. http://dx.doi.org/10.1017/S0140525X00035032

Thompson, R., & McConnell, J. (1955). Classical conditioning in the planarian, *Dugesia dorotocephala*. *Journal of Comparative and Physiological Psychology, 48,* 65–68. http://dx.doi.org/10.1037/h0041147

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23,* 87–102. http://dx.doi.org/10.3758/s13423-015-0892-6

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to change stimulus evaluations by means of subliminal approach and avoidance training. *Journal of Personality and Social Psychology, 110,* e1–e15. http://dx.doi.org/10.1037/pspa0000039

Weidemann, G., Best, E., Lee, J. C., & Lovibond, P. F. (2013). The role of contingency awareness in single-cue human eyeblink conditioning. *Learning & Memory, 20,* 363–366. http://dx.doi.org/10.1101/lm.029975.112

Weidemann, G., Satkunarajah, M., & Lovibond, P. F. (2016). I think, therefore eyeblink: The importance of contingency awareness in conditioning. *Psychological Science, 27,* 467–475. http://dx.doi.org/10.1177/0956797615625973

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science, 6,* 291–298. http://dx.doi.org/10.1177/1745691611406923

Appendix

Procedures and Stimuli for All Experiments

Table A1
Procedures for All Experiments

Experiment ID in article	Experiment ID in Supplement	Experiment ID in archive	CS–US interval	Acquisition CS duration	Conditioning test CS duration	Visibility test CS duration	Acquisition no. of trials	Conditioning test no. of trials	Visibility no. of test trials
1a	S1	3	75	75	25	25	240	144	144
	S2	2	75	75	25	25	240	288	288
1b	S3	4	75	75	25	25	120	72	72
1b	S4	4	75	25	75	75	120	72	72
1b	S5	4	75	25	25	25	120	72	72
	S6	17b	75	33	33	33	336	144	192
	S7	1	75	42	42	42	240	288	288
	S8	27	42[a]	42[a]	42[a]	42[a]	120	96	144
	S9	27	75	42	42	42	120	96	144
	S10	18a	75	58	58	58	336	192	192
	S11	23b	75	58	58	58	250	96	96
	S12	23c	75	58	58	58	250	96	96
1a	S13	3	75	75	75	75	240	144	144
1b	S14	4	75	75	75	75	120	72	72
	S15	17b	75	75	75	75	336	144	192
	S16	24US	75	75	75	75	250	96	96
	S17	24Belg.	75	75	75	75	250	96	96
1c	S18	26a+b+c	75	75	75	75	192	96	144
2a (1)		25a	375	375	375		250	96	
2a (2)		25b	375	375	375		250	96	
2a (3)		25c	375	83[b]	83[b]		250	96	
2b (1)		26a	375	375	375		168	96	
2b (2)		26b	375	75[c]	75[c]		168	96	
2b (3)		26c	375	75[d]	75[d]		168	96	

[a] Experiment S8: A 33-ms mask preceded the conditioned stimulus (CS), making the CS–unconditioned-stimulus (US) interval 42 ms, different from the 75-ms interval between CS onset and US in all other Series 1 experiments. [b] Experiment 2a(3): A 292-ms backward mask followed the 83-ms CS. Because there was no forward mask, the CS was visible. [c] Experiment 2b(2): A 300-ms blank screen followed the 75-ms CS. [d] Experiment 2b(3): A 100-ms backward mask, followed by 200-ms blank screen, followed the 75-ms CS.

(*Appendix continues*)

Experiment ID in article	Experiment ID in Supplement	Contingency awareness test	Experiment year	Total subject *N*	Retained subject *N*	US stimuli[h]	Counterbalancing	Advance contingency information
1a	S1	None	2007	180	97[i]	Gender (A, B)	j, k, n	
	S2	None	2006	78	47[i]	Gender (A, B)	j, k	
1b	S3	None	2007	93	85	Gender (A, B)	j, k, n	
1b	S4	None	2007	93	85	Gender (A, B)	j, k, n	
1b	S5	None	2007	93	85	Gender (A, B)	j, k, n	
	S6	None	2010	51	43	Valence (A)	j, k, n	
	S7	None	2006	113	73[i]	Gender (A, B)	j, k	
	S8	After CT	2015	75	67	Gender (A, B), Valence (A, B)	j, k, m, n	specific
	S9	After CT	2015	74	69	Gender (A, B), Valence (A, B)	j, k, m, n	specific
	S10	Before CT[f]	2010	104	92	Valence (A)	j, k	
	S11	During acquisition[g]	2011	56	53	Gender (A, B), Valence (A, B)	j, k, m	
	S12	During acquisition[g]	2012	89	82	Gender (A, B), Valence (A, B)	j, k, m	
1a	S13	None	2007	180	96[i]	Gender (A, B)	j, k, n	
1b	S14	None	2007	93	85	Gender (A, B)	j, k, n	
	S15	None	2010	64	51	Valence (A)	j, k, n	
	S16	None	2012	174	145	Gender (A, B), Valence (A, B)	j, k, l, m	
	S17	None	2013	32	32	Valence (A, B)	j, k, l, m	
1c	S18	After CT	2014	386	350	Gender (A, B), Valence (A, B)	j, k, l, m	non-specific[e]
2a (1)		Before CT	2013	81	66	Gender (A, B), Valence (A, B)	j, k, l, m	
2a (2)		Before CT	2013	58	46	Gender (A, B), Valence (A, B)	j, k, l, m	specific
2a (3)		Before CT	2013	109	101	Gender (A, B), Valence (A, B)	j, k, l, m	
2b (1)		After CT	2014	85	73	Gender (A, B), Valence (A, B)	j, k, l, m	non-specific
2b (2)		After CT	2014	99	92	Gender (A, B), Valence (A, B)	j, k, l, m	
2b (3)		After CT	2014	193	172	Gender (A, B), Valence (A, B)	j, k, l, m	

Note. CT = conditioning test.
[e] Experiment 1c: No contingency information was provided to the last 75% of Ss. Because the contingency information had no effect, all subjects were combined for reported analyses. [f] Experiment S10: The contingency awareness test was used for a special purpose, described in the General Discussion. [g] Experiment S11 and Experiment S12: The contingency awareness test was administered between the 3rd and 4th (of 5) acquisition trial blocks. [h] Letters in parentheses indicate sub-alphabets used in each experiment, and whether the experiment used only gender USs, only valence USs, or both. [i] High subject exclusion rates because of a programming error in the visibility test, which rendered that procedure's data unusable for a large fraction of subjects. [j] Counterbalancing of sub-alphabet used for CS vs US. [k] Counterbalanced assignments of specific letters within sub-alphabets to CSs associated with each US category. [l] Counterbalanced assignment of left key to female or unpleasant vs. male or pleasant. [m] Counterbalanced assignment of subjects to gender vs. valence USs. [n] CS durations in acquisition and/or test crossed with stimulus assignment variations.

Table A2
Stimulus Items for All Experiments

Experiment and data locations	Subalphabet A (*ADFHJLNOPWY*) CS consonant strings	Subalphabet A (*ADFHJLNOPWY*) US gendered names or valenced words	Subalphabet B (*BCEGIKMQRSTUVXZ*) CS consonant strings	Subalphabet B (*BCEGIKMQRSTUVXZ*) US gendered names or valenced words
1a (Figs. 2, S1, S13); also Figs. S2, S7	{*FJNP, JNPF, NPFJ, PFJN*} vs. {*DHLW, HLWD, LWDH, WDHL*}	female: *joan, donna, dawn, lola, polly, wanda, anna, hannah, holly, ann, nora, Wynona*; male: *john, jonah, andy, dan, alan, wally, noah, dylan, nolan, ladd, jay, donald*	{*BMVZ, MVZB, VZBM, ZBMV*} vs. {*GKQX, KQXG, QXGK, XGKQ*}	female: *eve, meg, susie, iris, tess, sue, vicki, bess, mimi, teri, keri, trixie*; male: *eric, curt, tim, mike, russ, kirk, merv, burt, steve, zeke, rick, emmet*
1b (Figs. 3, S3, S4, S5, S14)	{*FJJF, JFFJ, FJFJ, JFJF*} or {*NPPN, PNNP, NPNP, PNPN*} vs. {*DHHD, HDDH, DHDH, HDHD*} or {*LWWL,WLLW, LWLW, WLWL*}	female: *joan, donna, dawn, lola, polly, wanda, anna, hannah, holly, ann, nora, Wynona*; male: *john, jonah, andy, dan, alan, wally, noah, dylan, nolan, ladd, jay, donald*	{*BVVB, VBBV, BVBV, VBVB*} or {*MZZM, ZMMZ, MZMZ, ZMZM*} vs. {*GQQG, QGGQ, GQGQ, QGQG*} or {*KXXK,XKKX, KXKX, XKXK*}	female: *eve, meg, susie, iris, tess, sue, vicki, bess, mimi, teri, keri, trixie*; male: *eric, curt, tim, mike, russ, kirk, merv, burt, steve, zeke, rick, emmet*
1c (Figs. 5, S18); **2a** (Fig. 6); **2b** (Fig..7); also Figs. S8, S9, S11, S12, S16	*NPLW* vs. *DHJF*	negative: *flaw, folly, flood, oaf, flop, old*; positive: *dandy, joy, play, happy, jolly, fond*; female: *joan, donna, lola, wanda, hannah, ann*; male: *john, dan, alan, wally, noah, donald*	*BMVZ* vs. *GKQX*	negative: *miser, sick, bitter, bruise, crime, victim*; positive: *kiss, trust, true, cute, best, better*; female: *meg, susie, tess, sue, vicki, bess*; male: *eric, tim, mike, kirk, burt, steve*
Figs. S6, S15		negative: *down, flaw, pall, jalopy, folly, flood, oaf, flop, plod, howl, old, wallop*; positive: *dandy, joy, play, happy, halo, pal, doll, dawn, opal, jolly, fond, panda*	*BMVZ* vs. *GKQX* or {*BMVZ, MVZB, VZBM, ZBMV*} vs. {*GKQX, KQXG, QXGK, XGKQ*}	
Fig. S10		negative: *down, flaw, pall, jalopy, folly, flood, oaf, flop, plod, howl, old, wallop*; positive: *dandy, joy, play, happy, halo, pal, doll, dawn, opal, jolly, fond, panda*	*BMVZ* vs. *GKQX*	
Fig. S17	*NPLW* vs. *DHJF*	negative: *dood, haat, fataal, laf, waan, dof*; positive: *loyal, hoop, lof, top, loon, fan*	*BMVZ* vs. *GKQX*	negative: *misbruik, ziek, virus, ruzie, smerig, crisis*; positive: *kus, succes, reis, muziek, zege, vers*

Note. CS = conditioned stimulus; US = unconditioned stimulus.

Received September 15, 2016
Revision received August 16, 2017
Accepted August 16, 2017 ■