Final version at Social and Personality Psychological Science

Beckes, L. Simons, K., Lewis, D., Le, A., & Edwards, W. L. (2017). Desperately seeking support: Negative reinforcement schedules in the formation of adult attachment associations. *Social Psychological and Personality Science*, 8, 229-238.

**Desperately Seeking Support: Negative Reinforcement Schedules in the Formation
of Adult Attachment Associations**

Lane Beckes¹, Kailey Simons², Danielle Lewis¹, Anthony Le¹, and Weston Edwards¹

1 - Bradley University Department of Psychology, 2 – Illinois State University

Department of Psychology

Author Note:

Correspondence concerning this article should be addressed to:

Lane Beckes

Department of Psychology

Bradley University

1501 W Bradley Ave.

Peoria, IL 61625

Email: lbeckes@bradley.edu

Ph# 309-677-4185

Word Count: 4,999

Author Contributions:

Beckes developed the hypotheses and study design. Simons, Lewis, Le, and Edwards collected data under the supervision of Beckes. All authors assisted in data analysis led by Beckes. All authors contributed sections to the first draft of the paper. Beckes edited the final draft of the paper. All authors have approved the final manuscript for submission.

Acknowledgements

We thank Dallas Garrison, Casey Grage, Wendy Hasse, Brett Hoffman, Kylie McKinney, Victor Medina, Marina Palakeel, and Clara Tostovarsnik for their help in collecting data.

Abstract

Much is unknown about adult attachment style formation. We investigate whether negative reinforcement schedules promote hallmark features of secure and anxious attachment styles in a shock-threat support-seeking paradigm. Participants ostensibly asked for help from another participant seated in another room. Each time a shock threat signal appeared they were to press a button to indicate their need for help. The supporter could then stop the imminent shock. The reliability of the supporters was varied such that some supporters were consistent (continuous reinforcement) whereas others were inconsistent (variable-ratio reinforcement). Results indicated that inconsistently responsive others, reinforcing on a variable-ratio schedule, led to heightened approach related attentional biases toward the supporter, measured by event related potentials, increased positive attachment associations with the supporter, implicitly measured via a lexical decision task, and more negative explicit evaluations of the supporter.

Key Words: Attachment, Emotion, Social Cognition, Interpersonal Processes, Person Perception

Desperately Seeking Support: Negative Reinforcement Schedules in the Formation of Adult Attachment Associations

The need for social connection permeates human life (Baumeister & Leary, 1995; Beckes & Coan, 2011; Bowlby, 1969/1982, Holt-Lunstad, Smith, & Layton, 2010). John Bowlby's attachment theory may be the most extensive and durable theory of human connection. Originally focused on child-parent relationships, attachment theory has been subsequently extended to adult relationships (Hazan & Shaver, 1987). Adult attachments come in many types and most theorists argue for two dimensions of attachment quality, anxiety and avoidance (Mikulincer, Shaver, & Pereg, 2003). Typically scholars classify individuals into three categories of attachment: secure, anxious, or avoidant (c.f., Fraley, Waller, & Brennan, 2000). Secure individuals comfortably seek support, and have optimistic views of relationships and social support. Anxious individuals tend to express fear of abandonment, ambivalence about support, clingy behavior, and pessimism about receiving support. Avoidant individuals, alternatively, tend to discount their need for support and avoid dependence on others.

Adult attachment styles are linked to earlier attachment dynamics from infancy through adolescence (e.g., Simpson, Collins, Tran, & Hayden, 2007) and have stable elements over time (Fraley, Vicary, Brumbaugh, & Roisman, 2011), but also vary as a function of specific relationships (Klohn, Weller, Luo, & Choe, 2005). Whereas the stable elements predict many outcomes, including the development of relationship specific attachment styles, several relationship outcomes are primarily linked to relationship specific attachment styles, which can differ from dispositional styles. Yet, little is known about how relationship specific attachment styles develop. Here we test

whether the psychological features of different styles can be promoted in novel relationships through conditioning by manipulating negative reinforcement schedules (Beckes and Coan, 2015). If distinct negative reinforcement schedules lead to response patterns related to disparate attachment styles, then relationship specific attachment styles may emerge from these basic learning processes through repeated iteration. Examining these processes could provide a foundation for linking mechanisms to descriptive data and provide some experimental data to compare to extant models of attachment formation (Beckes & Coan, 2015; Mikulincer, et al., 2003).

The Origins of Individual Differences in Attachment

Mikulincer and Shaver (2005) argue for a support seeking dynamic in which individuals learn to use different strategies to manage distress, which normatively promotes proximity seeking to attachment figures. If the attachment figure is responsive to that distress, then the individual experiences well-being and safety (secure pattern). If the attachment figure is inalterably unresponsive, then the individual will employ deactivating strategies, distancing oneself from threat and attachment-related cues (avoidant pattern). If the attachment figure is unresponsive, but can be coaxed into providing support, the individual will employ hyperactivating strategies, such as clinginess, chronic attachment system activation, and experience ambivalent emotional responses (anxious pattern). When repeated over time, the individual will develop a *style* within the relationship.

Beckes and colleagues (Beckes & Coan, 2013, 2014, 2015; Beckes, Coan, & Morris, 2013; Beckes, Ijzerman, & Tops, 2015; Beckes Simpson, & Erickson, 2010) argue that the formation of attachment styles can be understood as a negative

reinforcement learning process with social responsiveness acting as an unconditioned reward (c.f., Harlowe, 1958). From this perspective, continuous negative reinforcement (consistent responsiveness) should produce secure patterns. Such consistency produces predictable contingencies between behavior and outcomes. As a result, support seeking occurs when needed, but not when unneeded, evaluations of consistent supporters become unambiguously positive and approach motivation toward that provider is enhanced.

Alternatively, variable-ratio schedules (inconsistent responsiveness) should produce ambivalent attachment attitudes and hyperactivation, such as chronic support seeking and strong approach motivation. This matches with long established findings that variable-ratio reinforcement produces steep response slopes in animal learning (Skinner, 1956), and increases the incentive salience of rewards (Anselme, Robinson, & Berridge, 2013). For example, unpredictable food rewards increase rats' approach speed, effort to achieve the reward, and vigor of engagement with conditioned cues.

Conditioning and Attachment Attitudes

In line with this approach, Beckes and colleagues (2010, 2013) found that repeatedly presenting a threat followed by a warm, smiling face led the face to prime the recognition of secure attachment related words and to inhibit the recognition of insecure attachment related words. This suggests that continuous negative reinforcement promotes positive implicit associations. Similarly, inconsistent support should be associated with relief, calm, etc., just unpredictably, and that unpredictability should increase the incentive salience (Anselme, Robinson, & Berridge, 2013) of that support, which should

increase the strength of the association between the inconsistent person and implicit security constructs. But what about implicit insecurity constructs?

Dispositionally, insecure individuals tend to have attachment representations of both a negative and positive valence chronically accessible (e.g., Mikulincer, Gillath, & Shaver, 2002; Zayas & Shoda, 2005; 2014). One possibility is that variable-ratio conditioning immediately produces associations like those seen in dispositional attachment anxiety, the dispositional hypothesis. But will those dispositional findings translate to the very first moments of learning within a relationship? It is plausible that conditioning primes positive attachment constructs and inhibits negative attachment constructs primarily through attachment motivation being satisfied via reinforcement, and does not immediately produce accessibility to bivalent attachment constructs. Thus, the associations may be a function of attachment motivation and reinforcement, not the manifestation of dispositional attitudes (the motivation and reinforcement hypothesis).

Either way, inconsistent responsiveness should promote explicit evaluations of supporters as untrustworthy and unpredictable, whereas consistent responsiveness should be viewed as trustworthy and predictable. Thus, inconsistency should produce ambivalence whether the dispositional or motivation and reinforcement predictions are correct. If the motivation and reinforcement hypothesis is correct, it seems likely the tension between intense approach motivation and explicit distrust ultimately promotes dispositional bivalent attachment associations seen in anxious attachment.

Conditioning and Approach Motivation

Evidence for increased approach motivation toward supporters after negative reinforcement comes from electroencephalography (EEG) and event related potential

(ERP) studies. In particular, the P1 potential, a positive deflection peaking around 100ms after face presentations, is associated with emotional attention biases. This component emerges early in processing and is associated primarily with attention modulation via external stimuli. Cunningham and colleagues (Cunningham, Van Bavel, Arbuckle, Packer, & Waggoner, 2012) suggest that social approach motivation modifies P1 responses. P1s are greater to faces toward which participants have an intrinsic approach bias, and increasing approach, as opposed to avoidance, motivation increases P1s. As such, the P1 may be useful in investigating whether negative reinforcement in attachment contexts increases P1 responses to conditioned faces given that the conditioning procedure is designed to promote the approach behavior of support seeking. Supportive of these predictions, Beckes, Coan, and Morris (2013) found increased P1 responses to continuously negatively reinforced faces indicating increases in approach-related attentional biases. But what about variable-ratio schedules? P1s should be greatest to unpredictable supporters given that unpredictable reinforcement increases approach motivation.

Experimental Design and Predictions

We employed a shock-threat conditioning paradigm, during which participants asked for help from a “supporter” to prevent being shocked. We manipulated, within-subjects, the reinforcement schedule such that some supporters were consistently helpful (continuous reinforcement), inconsistently helpful (variable-ratio reinforcement), and novel at test (neutral). We had three primary dependent variables (DVs), implicit attachment attitudes toward each face (**response times** in a lexical decision task to secure and insecure attachment words), **explicit attitudes** toward each face, and approach-

related attentional biases (**P1** ERP component) to each face. We chose these DVs because they allowed us to test for attitudinal ambivalence and approach-related attention.

During test we repeatedly presented each face as a prime in a lexical decision task. To measure both positive and negative implicit attachment associations we employed secure attachment words and insecure attachment words. In addition we measured non-attachment words to ascertain whether learning was restricted to the attachment domain. The lexical decision task was a 3 *Conditioning* schedule (continuous vs. variable-ratio vs. neutral) X 2 *Wordtype* (attachment vs. non-attachment words) X 2 *Valence* (positive vs. negative words) within-subjects design. For the P1 responses we tested left and right hemisphere electrodes resulting in a 3 *Conditioning* X 2 electrode *Hemisphere* within-subjects design. Finally we measured explicit ratings of each face, leading to a 3 level *Conditioning* within-subjects design.

For each DV we had a priori hypotheses (see OSF website for registered predictions, materials, design, and code: https://osf.io/erk9c/?view_only=de968b5570884016afbfc5674a52c3d7, preregistration materials were consolidated before any data were processed, but shortly *after* data collection had started), see Table 1 for predictions and outcomes. For response times (RTs) we predicted a three-way interaction between *Conditioning*, *Wordtype*, and *Valence*, driven by an interaction between *Conditioning* and *Valence* in the attachment words, with no effects on the non-attachment words. We then produced paired-comparison predictions. For the variable-ratio condition we predicted primes would facilitate recognition of secure words: variable-ratio primed secure word RTs < continuous primed secure word RTs, 2) variable-ratio primed secure word RTs < neutral

primed secure word RTs. For insecure words the dispositional hypothesis would predict: 1) variable-ratio primed insecure word RTs < continuous primed insecure word RTs; 2) variable-ratio primed insecure word RTs < neutral primed insecure word RTs, whereas the motivation and reinforcement hypothesis would predict: 1) variable-ratio primed insecure word RTs > continuous primed insecure word RTs; 2) variable-ratio primed insecure word RTs > neutral primed insecure word RTs. For the continuous conditioning we predicted (see Beckes, et al., 2010; 2013): 1) continuous primed secure word RTs < neutral primed secure word RTs, 2) continuous primed insecure word RTs > neutral primed insecure word RTs.

For the P1 ERP, we predicted a significant Conditioning X Hemisphere interaction, and a main effect of Conditioning on P1s. Specifically, variable-ratio conditioned faces would produce the largest P1 responses, continuously conditioned faces the second largest P1 responses, and neutral faces the smallest P1 responses. Further, we expected this pattern to be strongest in the left hemisphere as found in Beckes et al. (2013).

Finally, it was predicted that variable ratio-faces would be explicitly evaluated more negatively than neutral faces due to their unpredictability and unreliability, and continuously reinforced faces would be evaluated more positively than neutral faces.

The pattern of results we predict from the variable-ratio schedule would show the type of motivational hyperactivation and ambivalence one might expect within the early development of a new relationship specific style when the new partner is inconsistently responsive. Alternatively, the type of moderate approach motivation and unambiguous

positive attitudes predicted to emerge out of the continuous schedule matches what we would expect of an early developing secure relationship.

Methods

Participants

Sixty participants (41 women), 18-43 years old ($M=19.5$), participated. Sixty-five percent identified as White, 8% African-American, 10% Asian, 12% Hispanic, and 3% other. Two participants were left hand dominant, and 4 spoke English as a non-primary language. Participants were compensated (extra credit or cash) for participation, and treated in accordance with all IRB requirements. We invoked a stop rule at 60 participants with complete data. This rule was based on a sample size estimation using G*Power paired t-test calculator (all manipulations were within-subjects) assuming a modest effect size of $d_z=.4$ and power of .80, needing around sixty participants¹.

Procedure

Set-up. After informed consent, we delivered a cover story that participants were in a study of the provision and receipt of social support with three other participants. We told them that they would at times act as a supporter of another participant and would be able to prevent the other person from getting an electric shock, and at other times would be supported by another participant who could prevent them from getting a shock. We took a digital photo of the participant (immediately erased) to prevent participant suspicion upon seeing photos of “other participants” in a later task. Subsequently they completed demographics questions, the Experience in Close Relationships Revised

¹ We used the paired t-test calculation as it was more conservative ($N=52$) than the within-subjects ANOVA F-test calculator after setting $\eta_p^2=.04$ corrected to include the intercorrelations of the data, number of observations to 12, and correcting for sphericity at .7 ($N=49$), but then added participants for design considerations.

(ECRR, attachment style measure), and researchers placed EEG electrodes on them. They then followed an Eprime practice script with instructions on the main tasks.

EEG preparation. EEG recordings included 30 scalp electrodes measured via the BrainVision ActiChamp battery-powered system. We used a headcap with electrode sites in extended international 10/20 locations (Fp1, Fp2, F3, F7, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8 O1, OZ, O2, TP9, TP10). Fpz served as the ground, and Cz as reference. SuperVisc 1000gram electrode gel reduced impedance between the scalp and Ag-AgCl electrodes.

Supportive task. After set up, participants experienced the supportive task. We designed this task to increase the believability of subsequent procedures by having the participant act as a supporter before being supported. A math sheet was given for the participant to complete and they were told to watch for signals over the computer (a red box) while completing the math problems. The math task was designed to provide a plausible reason that someone might miss a signal for help from another participant. If they saw a signal, they were told they could prevent a shock from being delivered to the “other participant” by pressing a button-box button. Details are in the supplementary online materials.

Learning task. During learning, participants saw red X’s, indicating imminent shock if the “supporter” failed to help, and blue O’s, indicating safety. A red X meant they should push a button-box button as frequently as needed to signal to a “supporter” in another room. This signaled the “supporter” to push a button to prevent the shock, see Figure 1. This task had two blocks. In one block the “supporter” was consistently

responsive (continuous reinforcement). In the other block a different “supporter” was inconsistently responsive (variable-ratio reinforcement). Block order and the block-“supporter” image pairings were completely counterbalanced across participants. Details of the instructions, stimulus presentation times, and supporter response profiles are available in the supplementary online materials.

Fig. 1

Depiction of the sequence of events during the learning task for both the inconsistent variable ratio learning condition and the consistent continuous reinforcement condition. Each trial begins with the presentation of an “X” indicating threat or “O” indicating safety. Participants are then allowed to push a button on a button box to ostensibly ask for help from another participant. In the variable ratio condition the other responds with varying speed in 75% of trials, but fails to respond in 25% resulting in a shock. In the continuous condition the other responds quickly and in 100% of trials.

Lexical decision task. Test trials followed learning. Researchers recorded EEG data and the participant started the Lexical decision task. Upon receiving and

understanding the instructions (see Supplementary Materials), participants did three practice trials and began. Stimulus presentations had a black background with white text. Trials included a 500ms fixation point, followed by a 1000ms prime face, followed by a word or non-word until a word/non-word judgment was made. Participants pushed the first button on the button-box for words and second button for non-words as quickly as possible. Each word/non-word was paired with each face once and pairings randomly presented.

Final Tasks. Subsequently participants completed SAM scales measuring affective arousal and valence during shock, and explicit ratings of each face (face-item pairings randomized). We then asked the participant to speculate on the predictions of the study, if anything appeared odd, and if anything seemed suspicious (suspicion was noted in the log notes). Finally, the participant was debriefed and thanked.

Materials

Self Assessment Manikin (SAM) scales. Self-Assessment Manikins (SAM) are a visual self-assessment self-report measure of valence and arousal (Bradley & Lang, 1994). Scales depict five anthropomorphic manikins as a visual Likert-scale (e.g., valence includes a broadly smiling figure for very positive and a deeply frowning figure on the for very negative). SAM scales were used as a manipulation check to ensure effectiveness of the threat.

Experiences in Close Relationships Revised (ECRR). The ECRR is a self-report measure of attachment style originally developed by Brennan, Clark, and Shaver (1998), and revised by Fraley, Waller, & Brennan (2000). It includes 18 anxiety items (e.g., “I often worry that my partner will not want to stay with me”) and 18 avoidance

items (e.g., “I find it difficult to allow myself to depend on romantic partners”), each rated on a seven-point Likert-scale. The ECRR exhibited high internal consistency reliability (avoidance $\alpha=.90$; anxiety $\alpha=.92$).

Explicit Rating Scale. Participants rated faces on eight positively valenced adjective items: trustworthy, predictable, kind, responsive, thoughtful, reliable, likable, and attractive. The rating scale involved agreeing or disagreeing with the degree to which each face possessed each quality on a 1-5 scale (1=strongly disagree, 5=strongly agree). Internal consistency reliability was high for faces in each conditioning group (Variable $\alpha=.91$; Continuous $\alpha=.93$; Neutral $\alpha=.87$).

Face Stimuli. Face stimuli were taken from the FEI face database (<http://fei.edu.br/~cet/facedatabase.html>). Three smiling female faces were used in this study (15-12.jpg, 79-12.jpg, and 99-12.jpg). Practice items included an unrelated male face.

Lexical Decision Word Lists. The lexical decision task consisted of the presentation of positive attachment words (e.g. secure, nurture), negative attachment words (e.g., avoid, cold), positive non-attachment words (e.g., lust, intercourse), negative non-attachment words (e.g., disease, rot), neutral words (e.g., gallon, tissue), and finally non-words (e.g., buntorn, joov). These words were taken from previous research by Beckes, et al. (2010) and Beckes, et al. (2013). Readers can refer to the OSF website for all word lists https://osf.io/erk9c/?view_only=de968b5570884016afbfc5674a52c3d7.

Math distracter. The math distracter consisted of twelve problems (e.g., If $f(x)=3x^3-7x+5$, then $f(-1)=$). This task was designed to be too difficult (some unsolvable) to complete in the time given.

Analysis and Data Reduction

Response time data. RTs were averaged per participant in each of the lexical decision conditions after removing outliers (below 300ms and 3SD above the mean, consistent with Beckes et al., 2013) and trials with incorrect responses (average accuracy was high at 92%). Data was analyzed in SPSS using repeated measures analysis of variance (rmANOVA) and further interrogated with rmANOVA and paired t-tests.

ERP analysis and data reduction. EEG was sampled at 500Hz, processed in EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Markley, Luck, & Lopez-Calderon, 2011). Data were filtered offline with an IIR Butterworth .1Hz high-pass filter and a 30Hz low-pass filter. An average reference was derived offline. Electrodes placed below the left and the right of the right eye captured eye movements and blink artifacts. Epochs were defined as 200ms pre- to 800ms post-stimulus presentation. ERPLAB's moving window peak-to-peak artifact detection function detected artifacts (voltage threshold of 80 μ v). Visual inspection verified this cutoff. Data were corrected to the mean voltage of the prestimulus interval. Eight participants failed to have at least 20 accepted trials for each face stimulus and were removed from analysis. Mean P1 amplitudes were extracted in each target electrode between 110–140ms after stimulus onset. Decisions on filters, average referencing, artifact detection, and the use the average amplitude between 110–140 ms post-stimulus presentation were identical to those used in Beckes et al. (2013). Specific analysis plans are documented on the OSF page https://osf.io/erk9c/?view_only=de968b5570884016afbfc5674a52c3d7 in the analyses and predictions document.

Results

Manipulation Check SAM Scales

One sample t-tests were performed on the SAM Scale data. We tested the mean of the sample against values that would represent no arousal (5) and neutral valence (3).

Arousal was significantly increased by shock (5=no arousal, 1=highest arousal), $t(59)=-14.00$, $p<.001$, 95% CI [-2.61, -1.96], ($M=2.72$, $SD=1.26$), indicating moderate to high arousal. Valence had a test value of 3, which indicates neutral valence (5=most positive, 1=most negative). Mean valence was significantly lower than 3, $t(59)=-6.42$, $p<.001$, 95% CI [-.94, -.49], ($M=2.28$, $SD=.87$), indicating a negative response to the shock.

Table 1

Predictions at each level of analysis for each dependent variable, fit with data by standard statistical cut-offs, fit with data by general pattern, number of tails for the test (applicable only to paired t-tests), and p-value of the resultant test.

Test Type	Specific Prediction	Fit Prediction by p-Cutoff	Fit Prediction by Pattern	Tails ²	P-Value
Response Times					
Omnibus Three-Way Interaction	Conditioning X Word Type X Valence Interaction	Yes	Yes	N/A	0.031
Nested Two-Way Interactions					
Attachment Words	Conditioning X Valence Interaction	Yes	Yes	N/A	0.028
Non-Attachment Words	No Conditioning X Valence Interaction (NULL)	Yes	Yes	N/A	0.506
Nested Paired-Comparisons					
Positive Attachment Words	Variable-Ratio < Continuous	No	Yes	1	0.06
	Variable-Ratio < Neutral	Yes	Yes	1	0.0025
	Continuous < Neutral	No	Yes	1	0.108

² One-tailed tests were used for specific directional hypotheses as is recommended by statistical experts (e.g., Cho & Abe, 2013). Given the a-prior nature of the hypotheses, and their empirical and theoretical basis, we believe that this approach was the most principled and appropriate.

Negative Attachment Words	Variable-Ratio $\neq$ Continuous	No	?	2	0.48
	Variable-Ratio $\neq$ Neutral	No	?	2	0.417
	Continuous > Neutral	No	?	1	0.464
Explicit Evaluations					
Omnibus ANOVA	Conditioning Main Effect	Yes	Yes	N/A	< .001
Nested Paired-Comparisons	Variable-Ratio < Continuous	Yes	Yes	1	< .001
	Variable-Ratio < Neutral	Yes	Yes	1	< .001
	Continuous > Neutral	Yes	Yes	1	< .001
P1 ERPs					
Omnibus ANOVA	Conditioning Main Effect	Yes	Yes	N/A	0.041
	Conditioning X Hemisphere Interaction	No	?	N/A	0.607
Nested Paired-Comparisons	Variable-Ratio > Continuous	Yes	Yes	1	0.035
	Variable-Ratio > Neutral	Yes	Yes	1	0.0085
	Continuous > Neutral	No	Yes	1	0.237

Lexical Decision Task

The lexical decision task involved priming words of different categories (word-type) and valence with conditioned faces. We predicted a three-way interaction between Conditioning of the priming face (variable-ratio vs. continuous vs. neutral), Wordtype (attachment vs. non-attachment), and word Valence (positive vs. negative). Significant results emerged for the main effect of Wordtype, $F(1,59)=33.32$, $p<.001$, $\eta_p^2=.361$, with faster response times to attachment words ($M=601.45$, $SE_3=1.63$) than non-attachment words ($M=620.26$, $SE=1.64$), 95% CI [-25.33, -12.29]. Finally the predicted interaction between Conditioning by Wordtype by Valence was significant, $F(1.98,116.96)=3.61$, $p=.03$. All standard errors reported are computed using the within-subjects standard error approach recommended by Cousineau (2005).

$p=.031$, $\eta_p^2=.058$,. See Figure 2 and Table 2 for descriptive statistics. No main effect of Conditioning was detected, $F(1.80,106.13)=.69$, $p=.489$, $\eta_p^2=.012$, no main effect of Valence was detected, $F(1,59)=3.83$, $p=.055$, $\eta_p^2=.061$, nor were any other interactions: Conditioning X Wordtype, $F(1.88,110.96)=.70$, $p=.492$, $\eta_p^2=.012$; Conditioning X Valence, $F(1.97,116.33)=.26$, $p=.769$, $\eta_p^2=.004$; Wordtype X Conditioning, $F(1,59)=.40$, $p=.531$, $\eta_p^2=.007$.

We further interrogated these effects to make paired-comparisons and test nested models. For word space reasons, these sub-analyses are in the supplementary online materials. These analyses indicated that the interaction between Conditioning and Valence was significant for the attachment words, but not the non-attachment words, comporting with predictions. Further analyses indicated a pattern of RTs to secure words consistent with predictions, but only the variable-ratio vs. neutral comparison met threshold. We found no significant effect for the insecure words. See Table 1 for a summary of these results.

Table 2

Means and within-subjects standard errors or lexical decision response times by condition.

Schedule	Attachment Words				Non-Attachment Words			
	Positive		Negative		Positive		Negative	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Variable-Ratio	587.78	3.69	609.36	5.25	619.14	6.02	616.03	6.47
Continuous	597.84	4.69	604.32	3.80	620.76	6.20	620.45	6.15

Neutral	605.63	5.28	603.78	4.22	613.92	4.28	623.26	6.81
---------	--------	------	--------	------	--------	------	--------	------

Fig. 2

Mean and within-subjects 95% confidence interval of response times in milliseconds to words primed by faces within each condition of the design. Face primes included the faces negatively reinforced by variable-ratio or continuous primes in the learning trial, and a novel neutral face. Types of words are split up by valence and word type (attachment related positive: secure words, attachment related negative: insecure words,

non-attachment positive: sexual desire words, and non-attachment negative: disgust words).

Explicit Evaluation

For explicit evaluations we predicted a main effect of Conditioning such that the variable-ratio condition would result in the most negative evaluations and the continuous condition would result in the most positive evaluations. A 3 factor within-in subjects ANOVA revealed a significant difference in explicit evaluations as a function of conditioning, $F(1.56, 92.18)=33.97, p<.001, \eta_p^2=.365$. As predicted, the continuous reinforcement faces ($M=4.05, SE=.09$) were rated most positively, followed by the neutral faces ($M=3.45, SE=.05$), and finally the variable-ratio faces ($M=2.95, SE=.09$). All pairwise comparisons were significant: variable-ratio vs. continuous, $t(59)=-6.66, p<.001, 90\% \text{ CI } [-1.38, -.82]$; variable-ratio vs. neutral, $t(59)=-4.47, p<.001, 90\% \text{ CI } [-.69, -.32]$; continuous vs. neutral, $t(59)=5.12, p<.001, 90\% \text{ CI } [.40, .79]$.

Fig. 3

Mean and within-subjects 95% confidence interval of mean explicit evaluations (A) on a 5-point likert-scale within each condition of the design (variable-ratio, continuous, and neutral), and the mean response to each individual ratings adjective (B) within each condition of the design.

ERP P1

Electrodes O1, O2, P3, and P4 were selected for analysis based on previous research (Beckes, Coan, & Morris, 2013). We conducted repeated-measures ANOVA with Conditioning (variable-ratio vs. continuous vs. neutral) and Hemisphere (left vs. right) as the IVs. We predicted a main effect of Conditioning such that variable-ratio face P1s would be greater than P1s to continuous faces, which would be greater than P1s to the neutral faces. Moreover, we predicted an interaction between Conditioning and

Hemisphere with an enhanced main effect in left hemisphere electrodes. A significant Conditioning main effect emerged, $F(1.98, 98.94)=3.31$, $p=.041$, $\eta_p^2=.062$. The pattern of results conformed to predictions, and two pairwise comparisons reached statistical significance. The variable-ratio conditioned faces ($M=5.63$, $SE=.10$) led to significantly greater activity than the continuously reinforced faces ($M=5.31$, $SE=.11$), $p=.07$, $p=.035$ one-tailed, 90% CI [.031, .618]. The variable-ratio conditioned faces led to significantly greater activity than the neutral faces ($M=5.17$, $SE=.11$), $p=.017$, $p=.0085$ one-tailed, 90% CI [.149, .776]. The continuously conditioned faces were non-significantly greater in activity than the neutral faces, $p=.473$, $p=.237$ one-tailed, 90% CI [-.182, .459], see Figure 4. The predicted interaction between Conditioning and Hemisphere did not emerge, $F(1.82, 91.02)=.48$, $p=.605$, $\eta_p^2=.009$.

Fig. 4

Depicts the mean and 95% confidence intervals of the P1 amplitude in the 110–140ms range (A) for electrodes P3, P4, O1, and O2 to faces that were wither novel (neutral) or negatively reinforced with a variable-ratio or continuous schedule, and average ERP response to faces in the O1 electrode (B) by condition.

Exploratory Analyses

Additional exploratory analyses were conducted, but only briefly mentioned here. We explored the relationship between the DVs as a function of condition. Only one significant effect emerged for the correlation between the continuous condition P1 difference and the continuous condition secure word speed, $r(51) = .321$, $p = .022$. In addition, given that general attachment styles are likely predictors in the development of relationship specific attachment styles, we found evidence of their influence through the ECRR. Moreover we explored the impact of gender on our conditioning data. Additionally, to ensure that suspicious participants were not driving any of the effects found, several analyses were performed with suspicion included as an independent variable. Suspicion was not a significant predictor in the lexical decision or ERP analyses, all p 's $> .05$. Suspicion did significantly interact with conditioning on the explicit ratings, $F(1.55, 90.05) = 5.076$, $p = .014$, $\eta_p^2 = .080$, but the effects held whether the participants were suspicious, $F(1.351, 24.31) = 20.27$, $p < .001$, $\eta_p^2 = .530$, or not, $F(1.65, 65.98) = 17.80$, $p < .001$, $\eta_p^2 = .308$. All analyses are available in the supplementary online materials.

Discussion and Conclusion

This experiment sheds light on the initial stages of relationship specific attachment formation exploring the hypothesis that negative reinforcement schedules are an important ingredient in the development of relationship specific adult attachment styles. It was hypothesized that variable-ratio negative reinforcement schedules should produce attitudinal ambivalence and hyperactivated approach motivation toward a supporter, whereas continuous reinforcement schedules should produce unambiguous positive attitudes and moderate approach motivation toward a supporter. Here, variable-ratio schedules produced features of attachment anxiety as predicted, including intensified approach motivation and attitudinal ambivalence toward a novel supporter. All a priori predictions of paired-comparisons between the variable-ratio condition and neutral condition were confirmed, except for our prediction regarding the insecure word RTs. This may be due to sampling variation, something about the particular methodological details of this study, or could be due to a true null difference. Moreover, we also detected greater P1 responses to the variable-ratio faces than to the continuous faces, and less positive evaluations of the variable-ratio faces relative to the other two conditions. The P1 results indicate an increase in approach motivation as a function of variable-ratio conditioning, whereas the facilitation of secure word RTs and negative explicit evaluations indicate attitudinal ambivalence.

Although the variable-ratio findings are generally supportive of our model, caution should be taken in interpreting these results. First, other measures of approach motivation should be used to verify our interpretation of the P1 effects found here, as only converging operations and findings can provide evidence for a truly strong interpretation. Second, these effects are constrained to specific relationship attachment

strategies in the earliest stages of development, and a very particular type of interaction. The type of support being provided here is primarily instrumental, but attachment processes often play out in more emotionally supportive contexts, such as giving someone a hug after a bad day. As such these findings must be supplemented with more ecologically valid data before generalizing beyond this context. Alternatively the data is high in internal validity and allows causal inferences that may be difficult in naturalistic studies. Ideally future research will bridge this gap with ecologically valid experiments with enough control to have moderate external *and* internal validity.

The data are also consistent with the hypothesis that continuous schedules produce secure associations, with P1 responses and secure concept priming intermediate between controls and variable ratio conditioned faces, and more positive explicit evaluations than controls, but outside of the explicit ratings the effects were unconvincing and failed to reject the null hypothesis. To examine the degree to which the continuous conditioning data comport with past studies we included a Bayesian analysis in the supplementary materials that examines the RT data in conjunction with data from Beckes et al. 2010 and 2013 (see supplementary table 5 and 6 and supplementary figure 1). This indicates the current data modestly strengthens the evidence for an effect of continuous negative reinforcement on implicit attachment attitudes ($BF_{10} = 8.7$ up from $BF_{10} = 6.0$), and the BF_{10} indicates substantial evidence for the alternative vs. null hypothesis (Jeffereys, 1961). These results indicate a possible lack of power, likely due to two factors. First, sample size was restricted due to EEG costs. Future studies with less time and cost intensive measures should collect larger sample sizes. Second, some participants may not have perceived a real threat in the continuous condition given the predictability

of the consistent responder. Future studies could include an occasional shock that the partner cannot stop in order to maintain shock credibility. Future studies should also consider the inclusion of other measures of approach motivation, and the development of more precise standardized measurements.

This data should not be interpreted as evidence that global attachment styles lack stability, or that this method is automatically capable of modifying either global or already intact relationship specific attachment styles. Future research should attempt to find ways to explore these possibilities. Exploratory analyses of general attachment anxiety and avoidance (see the supplementary material for details) replicated several previous findings in the literature, such as anxious individuals responding generally faster to bivalent attachment constructs, avoidant individuals slower to respond to all constructs, and avoidant attachment altering P1 responses to emotional faces (Dan & Raz, 2012). These results indicate the possibility that anxious individuals respond to social stimuli with vigilance increasing response times, whereas avoidant individuals respond with either disfluency in processing social information, or an attempt to suppress automatic responses to social stimuli, leading to interference. Thus, these stable individual differences are clear in our data, and interacted in some cases, although ambiguously, with the conditioning procedure, (e.g., anxious individuals responded more quickly to all words in the variable-ratio condition).

Bridging a gap between behaviorist and attachment theorists that persisted for many decades, this study provides some of the first direct support that negative reinforcement processes can influence attachment attitudes and motivation. The quality of attitudes and motivations may emerge, at least in part, out of the schedule of

reinforcement in predictable ways consistent with extant models (Beckes & Coan, 2015; Mikulincer et al., 2001). These ideas must be tested in increasingly ecologically valid contexts before acceptance of any application value, but if it generalizes, such information could help couples with attachment related conflicts (c.f., Johnson et al., 2013) and be a boon to psychological research and practice (c.f., Mikulincer & Shaver, 2007).

References

- Anselme, O. Robinson, M. J. F., & Berridge, K. C. (2013). Reward uncertainty enhances incentive salience attribution. *Behavioral and Brain Research*, 238, 53-61.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497-529.
- Beckes, L., & Coan, J. A. (2011). Social baseline theory: The role of social proximity in emotion and economy of action. *Personality and Social Psychology Compass*, 5, 976-988.
- Beckes, L. & Coan, J. A. (2013). Toward an integrative neuroscience of relationships. In J. A. Simpson & L. Campbell (Eds.), *The Oxford Handbook of Close Relationships*, (pp. 684-710). New York: Oxford University Press.
- Beckes, L., & Coan, J. A. (2014). Relationship neuroscience. In J. A. Simpson & J. Dovidio (Eds.), *The Handbook of Personality and Social Psychology, Volume 3: Interpersonal Relationships and Group Processes*. Washington DC., American Psychological Association Press.
- Beckes, L., & Coan, J. A. (2015). The distress-relief dynamic in attachment bonds. In C. Hazan & V. Zayas (Eds.), *Bases of Adult Attachment: Linking Brain, Mind, and Behavior*.
- Beckes, L., Coan, J. A., & Morris, J. P. (2013). Implicit conditioning of faces via the social regulation of emotion: ERP evidence of early attentional biases for security conditioned faces. *Psychophysiology*, 50, 734-742.
- Beckes, L., Ijzerman, H. & Tops, M. (2015). Toward a radically embodied neuroscience

- of attachment and relationships. *Frontiers in Human Neuroscience*. Doi: 10.3389/fnhum.2015.00266
- Beckes, L., Simpson, J. A., & Erickson, A. B. (2010). Of snakes and succor: Learning secure attachment associations with novel faces via negative stimulus pairings. *Psychological Science*, 21, 721 – 728.
- Bowlby, J. (1969/1982). Attachment and loss, volume I: Attachment. New York, NY: Basic Books.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49-59. doi:10.1016/0005-7916(94)90063-9
- Cho, H-C. & Abe, S. (2013). Is two-tailed testing for directional research hypotheses legitimate? *Journal of Business Research*, 6, 1261-1266.
- Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, 1, 42–45.
- Cunningham, W. A., Van Bavel, J. J., Arbuckle, N. L., Packer, D. J., & Waggoner, A. S. (2012). Rapid social perception is flexible: Approach and avoidance motivational states shape P100 responses to other-race faces. *Frontiers in Human Neuroscience*, 6, doi: 10.3389/fnhum.2012.00140.
- Dan, O., & Raz, S. (2012). Adult attachment and emotional processing biases: An event related potentials (ERPs) study. *Biological Psychology*, 91, 212-220.
- Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-21.

- Fraley, R. C., Vicary, A. M., Brumbaugh, C. C., & Roisman, G. I. (2011). Patterns of stability in adult attachment: An empirical test of two models of continuity and change. *Journal of Personality and Social Psychology*, 101, 974-992.
- Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item-response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, 78, 350-365.
- Harlow, H. (1958). The nature of love. *American Psychologist*, 13, 673-685.
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 5, 511-524.
- Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta- review. *Plos Medicine*, 7, e1000316.
doi:10.1371/journal.pmed.1000316
- Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.
- Klohnen, E. C., Weller, J. A., Luo, S., & Chloe, M. (2005). Organization and predictive power of general and relationship-specific attachment models: One for all, and all for one? *Personality and Social Psychology Bulletin*, 31, 1665-1682.
- Markley, C., Luck, S., & Lopez-Calderon, J. (2011). ERPLAB toolbox user's manual. URL: <http://erpinfo.org/erplab/erplab-documentation/manual>, accessed June 2012.
- Mikuincer, M., Gillath, O., & Shaver, P. R. (2002). Activation of the attachment system in adulthood: Threat-related primes increase the accessibility of mental

- representations of attachment figures. *Journal of Personality and Social Psychology*, 83, 881-895.
- Mikulincer, M., & Shaver, P. R. (2005). Attachment theory and emotions in close relationships: Exploring the attachment-related dynamics of emotional reactions to relational events. *Personal Relationships*, 12, 149-168.
- Mikulincer, M., & Shaver, P.R. (2007). Boosting attachment security to promote mental health, pro-social values, and inter-group tolerance. *Psychological Inquiry*, 18, 139–165.
- Mikulincer, M., Shaver, P. R., & Pereg, D. (2003). Attachment Theory and Affect Regulation: The Dynamics, Development, and Cognitive Consequences of Attachment-Related Strategies. *Motivation and Emotion*, 27, 77-102.
- Simpson, J. A., Collins, W. A., Tran, S., & Haydon, K. C.. (2007). Attachment and the experience and expression of emotion in romantic relationships: A developmental perspective. *Journal of Personality and Social Psychology*, 92, 355-367.
- Skinner, B. F. (1956). A case history in scientific method. *American Psychologist*, 11, 221-233.
- Zayas, V. & Shoda, Y. (2005). Do automatic reactions elicited by thoughts of romantic partner, mother, and self relate to adult romantic attachment? *Personality and Social Psychology Bulletin*, 31, 1011-1025.
- Zayas, V. & Shoda, Y. (2014). Love you? Hate you? Maybe its both: Evidence that significant others trigger bivalent priming. *Social Psychological and Personality Science*, 6, 56-64.

Supplementary Materials

Supplementary Methods

Learning Task Instructions

During the following task you will be randomly shown red X's and blue O's on the screen. If you see a red X you might be shocked. To avoid being shocked press the first button on the button box to send a signal to the participant in the adjacent experimental room. Your signal will only be displayed briefly, so you may need to push the button several times before they respond. If they respond in time, you will see a photo of the other participant. If they do not respond in time you will see a red box and receive a brief shock to your fingers. Please discuss the instructions with the researcher and ask any questions you have at this time.

Lexical Decision Task Instructions

For the following tasks you will see a fixation point, followed by a face, followed by a letter string (repeatedly many times). Each time you see a letter string push the first button on your box if it is a word, and the second button on your button box if it is not a word. Make your choice as fast as possible while still being accurate. If you understand press any key to do three practice trials.

Supportive Task Details

The task involved solving as many math problems as possible, while still paying enough attention to the computer screen to notice when a red box flashed upon the screen. The block consisted of 64 trials, which either unfolded with no changes to a blank screen displayed for 5000ms (32 trials), or during which the blank screen alternated with a flashing red box (32 trials). Participants were instructed that they could prevent a shock to the other participant if they pushed the first button on the button box quickly after noticing a red box, and if they failed to do so, or chose not to do so then the other participant would receive a shock. After completion participants were told that it would be their turn to undergo the shock threat procedure (aka, learning task).

Learning Task Details

In the continuous reinforcement condition the supporter consistently responded to the signal for help within a window of .5 - 1 seconds after the first signal for help was sent. In the variable ratio reinforcement condition the supporter responded to the signal within a window of .5 - 1 seconds after the first signal in 1/4 of the trials, within a window of .5 - 1 seconds after the 5th signal in 1/4 of the trials, within a window of .5 - 1 seconds after the 10th signal in 1/4 of the trials, and not at all in 1/4 of the trials. The trials were set up to include 10 frames of the red X presentation, which appeared to simply be one

presentation. Each X frame would last for 1 second and move to the next stimulus, or automatically move if the participant pressed the button box. For each block, a total of 32 safety and 32 shock trials were presented. Each time the “supporter” responded, their picture would display for 500 ms, before starting the next round. If they did not, the participant would receive a brief shock lasting 1000 ms.

Analysis Details

Data from one participant had an accuracy rate below 70% (closer to chance than the sample mean) indicating inattention or not following instructions. This participant was included in primary tests despite this concern because we did not include this as an exclusion in the preregistration. Tests excluding this participant are included in the OSF results files. Results are nearly identical regardless of the inclusion or exclusion of the participant. Similar decisions were made regarding participants for whom English is not a primary language. Again, however, their inclusion/exclusion made little difference.

Supplementary Analyses

Interrogation of Conditioning by Wordtype by Valence interaction

Separate ANOVA models testing the effects of Conditioning and Valence were conducted on attachment and non-attachment words. Here we predicted a significant two-way interaction between Conditioning and Valence in the attachment words, but no significant interaction in the non-attachment words under the assumption that the attachment words would drive the higher-level interaction. No significant main effects of Conditioning for the non-attachment words emerged, $F(1.98, 116.82) = .62$, $p = .537$, $\eta_p^2 = .010$, or Valence, $F(1, 59) = .76$, $p = .388$, $\eta_p^2 = .013$, were detected. Supporting our prediction, the interaction of Conditioning and Valence was similarly non-significant, $F(1.92, 113.34) = .68$, $p = .506$, $\eta_p^2 = .011$. A Bayesian repeated-measures ANOVA done in JASP confirmed this inference indicating strong evidence for the null hypothesis over the alternative, $BF_{01} = 11.10$.

For the attachment words, no significant main effect of Conditioning, $F(1.879, 110.86) = .84$, $p = .429$, $\eta_p^2 = .014$, was detected. However a significant effect of Valence, $F(1, 59) = 4.77$, $p = .033$, $\eta_p^2 = .075$, was detected indicating faster response times to positive words ($M = 597.08$, $SE = 2.84$) than to negative words ($M = 605.82$, $SE = 2.29$), 95% CI [-16.74, -.73]. As predicted a significant interaction between Conditioning and Valence was detected, $F(1.99, 117.12) = 3.71$, $p = .028$, $\eta_p^2 = .059$. Further interrogating this interaction indicated the predicted pattern of results in response times to secure words via a significant conditioning effect, $F(2, 117.79) = 4.07$, $p = .020$, $\eta_p^2 = .065$. We specifically predicted that response times to secure words primed by the neutral face would be slower than those primed by either the variable-ratio or continuously conditioned faces. Pairwise comparisons indicated a non-significant difference in response times between the variable-ratio reinforced condition, ($M = 587.78$, $SE = 3.81$) and the continuously reinforced condition ($M = 597.84$, $SE = 4.60$), $p = .121$, $p = .06$ one-tailed, 90% CI [-20.76, .64], a significant difference between the variable-ratio reinforced condition and the

neutral condition ($M=605.63$, $SE=5.62$), $p = .005$, $p = .0025$ one-tailed, 90% CI [-28.19, 23.75], and a non-significant difference between the continuously reinforced and neutral conditions, $p = .216$, $p = .108$ one-tailed, 90% CI [-18.19, 2.61] trending in the predicted direction. The predicted effect on insecure words as a function of conditioning did not emerge, $F(1.92, 113.37)=.43$, $p=.644$, $\eta_p^2=.007$.

Correlations Between Dependent Measures as a Function of Condition

We conducted a series of correlational tests to see if any of the DVs correlated as a function of condition. To do so we calculated change scores for each DV using the neutral condition as the baseline. For response times we calculated variable-ratio condition speed scores by subtracting the secure word response times in the variable-ratio condition from the secure word response times in the neutral condition. We calculated continuous condition speed scores in the same manner. This created a variable that represents faster response times to secure words in the variable-ratio or continuous conditions with higher speed scores, so the higher the score the faster they were relative to baseline. To get a the P1 difference score we subtracted the neutral P1s from the variable-ratio and continuous P1s separately to produce a P1 change score with higher scores indicating greater P1s to the variable-ratio and continuous conditioned faces. The same method was used for the explicit rating, with the difference scores indicating more positive evaluations with higher scores. Only one significant effect emerged for the correlation between the continuous condition P1 difference and the continuous condition secure word speed, $r(51)=.321$, $p=.022$. This indicated that larger P1s were associated with faster response times to secure words in the continuous condition, which mirrors a similar finding in Beckes and colleagues (2013). The correlation between the variable-ratio P1 difference and variable-ratio secure word speed, however, was close to zero, $r(51)=-.05$, $p=.726$. There were trending correlations indicating that faster response times to secure words might be associated with more positive explicit evaluations in the same condition: variable-ratio, $r(60)=.18$, $p=.17$; continuous, $r(60)=.17$, $p=.18$. These results are insufficient to make any claims upon, but the full correlational table can be seen in the supplementary analysis file on the OSF website.

Effects of Gender and Attachment Style

Lexical Decision Task. In the lexical decision data we first conducted a repeated-measures ANOVA with Conditioning, Wordtype, and Valence as within subjects factors, Gender as a between subjects factor, and ECRR Avoidance and ECRR Anxiety as covariates. A significant interaction between Valence and Gender emerged, $F(1, 54)=5.92$, $p = .018$, $\eta_p^2 = .099$. Patterns indicate that women responded with very similar speeds to both positive ($M=591.76$, $SE=$) and negative words ($M=592.48$, $SE=$) whereas men were faster to respond to positive words ($M=618.84$, $SE=$) than they were to negative words ($M=637.30$, $SE=$). An interaction between Conditioning, Wordtype, and Gender was also detected, $F(1.92, 103.70)=5.43$, $p = .006$, $\eta_p^2 = .091$. Subsequent analysis of means and paired t-tests indicated this effect was largely driven by much faster response times by men to disgust words primed by variable-ratio conditioned faces,

particularly versus response times to disgust words primed by continuously conditioned faces, $t(18)=-2.41, p=.027$, 95% CI [-91.53, -6.28], see Supplementary Table 1. This effect dropped the men's mean response to variable-ratio non-attachment words leading to the interaction effect. The meaning of these effects is difficult to decipher, but both effects may be a function of the gender of the target faces interacting with the gender of the participant in such a way that engages sexual motivations. The former may indicate a tendency for men to have more positive associations with women, particularly in the sexual and attachment domains. The latter effect in which men have an implicit disgust association may indicate that men engage in an automatic derogation of women who respond unpredictably.

Supplementary Table 1

Means and Within-Subjects Standard Errors of Response Times for Each Condition in the Condition x Wordtype x Gender Interaction

		Women		Men	
		Mean	SE	Mean	SE
Variable	Attachment	576.47	4.19	623.86	6.09
	Non-Attachment	607.70	4.95	616.54	7.19
Continuous	Attachment	583.66	3.46	614.52	5.02
	Non-Attachment	601.78	5.83	649.58	8.46
Neutral	Attachment	582.20	4.06	629.52	5.89
	Non-Attachment	600.88	4.78	634.52	6.94

For the ECRR, a significant interaction between Conditioning, Wordtype, and Anxiety emerged, $F(1.92, 103.70)=3.50, p=.036, \eta_p^2=.061$, as did a main effect of Anxiety $F(1, 54)=5.26, p=.026, \eta_p^2=.089$, and a main effect of Avoidance, $F(1, 54)=6.53, p=.013, \eta_p^2=.108$. Subsequent analysis of zero order correlations indicated a general tendency for negative correlations between Anxiety and response time, indicating anxious participants were generally faster to respond to all words, but no correlations were individually significant. Zero order correlations of avoidance and response times indicated the opposite pattern, indicating slower response times to words generally. To further interrogate these effects we ran partial correlations of Anxiety with response times statistically controlling for Avoidance, and ran partial correlations of Avoidance with response times controlling for Anxiety, see Supplementary Tables 2 and 3. With the partial correlations it is evident that Anxiety is significantly negatively correlated with response time across all word categories in the variable-ratio condition, except the secure category, although it is marginal. In the continuous condition, this negative association is only significant in the attachment word categories, and in the neutral condition, the negative association is only significant with the non-attachment words (disgust in particular, with sexual desire marginal). This indicates that the general tendency of anxious individuals to respond quickly is enhanced in the variable-ratio condition. Alternatively, Avoidance is significantly positively associated with response times in almost every word category, and marginally in those that don't meet conventional thresholds. These effects may reflect stable individual differences in response to social

stimuli. The Avoidance effect may indicate disfluency in processing social information that interferes with the subsequent lexical decision. The Anxiety effect may indicate heightened attention is triggered by the social stimulus, producing faster response times in the subsequent lexical decision task.

Supplementary Table 2

Depiction of correlations between attachment anxiety and response times as a function of conditioning, wordtype, and valence, statistically controlling for attachment avoidance

Conditioning	Attachment Words		Non-Attachment Words	
	Secure	Insecure	Sexual Desire	Disgust
Variable-Ratio	-.226 <i>p</i> = .091	-.306 <i>p</i> = .021	-.305 <i>p</i> = .021	-.290 <i>p</i> = .029
Continuous	-.276 <i>p</i> = .038	-.333 <i>p</i> = .011	-.138 <i>p</i> = .304	-.164 <i>p</i> = .222
Neutral	-.182 <i>p</i> = .176	-.189 <i>p</i> = .159	-.239 <i>p</i> = .073	-.273 <i>p</i> = .040

Supplementary Table 3

Depiction of correlations between attachment avoidance and response times as a function of conditioning, wordtype, and valence, statistically controlling for attachment anxiety

Conditioning	Attachment Words		Non-Attachment Words	
	Secure	Insecure	Sexual Desire	Disgust
Variable-Ratio	.332 <i>p</i> = .012	.358 <i>p</i> = .006	.364 <i>p</i> = .005	.384 <i>p</i> = .003
Continuous	.263 <i>p</i> = .048	.415 <i>p</i> = .001	.249 <i>p</i> = .062	.260 <i>p</i> = .051
Neutral	.323 <i>p</i> = .014	.299 <i>p</i> = .024	.332 <i>p</i> = .012	.265 <i>p</i> = .046

Explicit Ratings. Attachment avoidance significantly interacted with conditioning to predict explicit ratings, $F(1.559, 84.18) = 4.06$, $p = .030$, $\eta_p^2 = .070$. A subsequent curve fit regression indicated a marginally significant positive correlation between Attachment Avoidance and explicit ratings in the Variable-Ratio condition, $r = .23$, $p = .078$, but no correlation in the Continuous, $r = -.09$, $p = .511$, or Neutral conditions, $r = -.10$, $p = .352$. This indicates that avoidant individuals rated inconsistent targets as more likeable in general. This may be due to a history of relationships with unresponsive others such that inconsistency is not seen as a particularly negative behavior.

P1 ERP Effects.

In the ERP data a significant main effect of Attachment Avoidance emerged, $F(1, 45)=8.01$, $p = .007$, $\eta_p^2 = .151$, indicating positive relationships between P1 responses and Avoidance. This is consistent with previous findings that avoidant individuals respond to novel faces with higher P1 responses, indicating early attentional biases (Dan & Raz, 2012). Analysis outputs can be seen at the OSF website, in the SupplementaryAnalyses.pdf file. An additional interaction between Condition x Avoidance also emerged, $F(1.96, 88.37)=4.33$, $p = .047$, $\eta_p^2 = .066$, but subsequent analysis indicated that this effect may be derived largely from model overfitting. There is a small decrease in overall correlation between avoidance and response times in the variable-ratio condition, but these correlations are still positive and significant.

Suspicion Effects

As noted in the primary document, suspicion did significantly interact with conditioning on the explicit ratings, $F(1.55, 90.05)=5.076$, $p = .014$, $\eta_p^2 = .080$, but the effects held whether the participants were suspicious, $F(1.351, 24.31)=20.27$, $p < .001$, $\eta_p^2 = .530$ or not, $F(1.65, 65.98)=17.80$, $p < .001$, $\eta_p^2 = .308$. See Supplementary Table 4 and Supplementary Figure 1 for details on the means, standard errors, and 95% confidence intervals for each condition. The pattern is identical in each group, but the suspicious participants rated the inconsistent (Variable-Ratio) target more negatively than did the non-suspicious group.

Supplementary Table 4

Means and Within-Subjects Standard Errors of Explicit Evaluation in Each Condition in the Condition x Suspicion Interaction

	Suspicious		Not Suspicious	
	Mean	SE	Mean	SE
Variable-Ratio	2.47	.18	3.17	.09
Continuous	4.11	.17	4.02	.10
Neutral	3.49	.08	3.44	.06

Supplementary Fig. 1

Depiction of Means and Within-Subjects 95% Confidence Intervals of Explicit Evaluation in Each Condition in the Condition x Suspicion Interaction

Bayesian Analysis of Continuous Conditioning Including Current Data and Data from Beckes and Colleagues (2010, 2013)

To examine whether the current data reduces confidence in the overall continuous conditioning approach we analyzed the data with the data from Beckes et al., 2010 and Beckes et al., 2013 using a Bayesian repeated-measures ANOVA in JASP (we included all participants from the 2013 paper, including those removed in the original paper). The $BF_{10} = 8.7$ indicates substantial support for the alternative hypothesis over the null hypothesis (Jeffreys, 1961), and the posterior probability was increased when the current data was included in the model (model without current data $BF_{10} = 6.045$), see supplementary tables 5 and 6. Our best current estimate of the effect is displayed in supplementary figure 1.

Supplementary Table 5

Bayesian analysis of continuous conditioning data including current data and data from Beckes et al., 2010 and Beckes, 2013.

Models	P(M)	P(M data)	BF _M	BF ₁₀	% error
Null model (incl. Conditioning, Valence, subject)	0.500	0.103	0.115	1.000	
Conditioning * Valence	0.500	0.897	8.700	8.700	5.392

Note. All models include Conditioning, Valence, subject.

Supplementary Table 6

Bayesian analysis of continuous conditioning data from Beckes et al., 2010 and Beckes, 2013 (excluding current data).

Models	P(M)	P(M data)	BF _M	BF ₁₀	% error
Null model (incl. Conditioning, Valence, subject)	0.500	0.142	0.165	1.000	
Conditioning * Valence	0.500	0.858	6.045	6.045	5.766

Note. All models include Conditioning, Valence, subject.

Supplementary Figure 1 shows the means and 95% confidence intervals of all continuous conditioning data from the current study, Beckes et al., 2010, and Beckes et al., 2013. Note that there is a general speed advantage when participants use a button box versus keyboard, and keyboard results should be slower than this estimate, while button box results should be faster.

References

- Beckes, L., Coan, J. A., & Morris, J. P. (2013). Implicit conditioning of faces via the social regulation of emotion: ERP evidence of early attentional biases for security conditioned faces. *Psychophysiology*, 50, 734-742.
- Beckes, L., Simpson, J. A., & Erickson, A. B. (2010). Of snakes and succor: Learning secure attachment associations with novel faces via negative stimulus pairings. *Psychological Science*, 21, 721 – 728.

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.