Cognitive Science 33 (2009) 1232–1251

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01054.x

Fear Conditioning and Social Groups: Statistics, Not Genetics

Tiago V. Maia

Department of Psychiatry, Columbia University

Received 19 January 2009; received in revised form 25 April 2009; accepted 6 May 2009

Abstract

Humans display more conditioned fear when the conditioned stimulus in a fear conditioning paradigm is a picture of an individual from another race than when it is a picture of an individual from their own race (Olsson, Ebert, Banaji, & Phelps, 2005). These results have been interpreted in terms of a genetic ''preparedness'' to learn to fear individuals from different social groups (Ohman, 2005; Olsson et al., 2005). However, the associability of conditioned stimuli is strongly influenced by prior exposure to those or similar stimuli. Using the Kalman filter, a normative statistical model, this article shows that superior fear conditioning to individuals from other groups is precisely what one would expect if participants perform optimal, Bayesian inference that takes their prior exposures to the different groups into account. There is therefore no need to postulate a genetic preparedness to learn to fear individuals from other races or social groups.

Keywords: Fear conditioning; Latent inhibition; Preparedness; Social groups; Kalman filter

1. Introduction

Olsson et al. (2005) showed that in a classical fear conditioning paradigm both Black and White participants displayed stronger and more enduring fear reactions when the conditioned stimulus (CS) was a picture of an individual from the other race than when it was a picture of an individual from their own race. The experiment used what is known as a differential conditioning paradigm (Ohman, Frederikson, Hugdahl, & Rimmo, 1976; Ohman & Mineka, 2001). The CSs were two pictures of Black individuals and two pictures of White individuals. The experiment consisted of three phases—habituation, acquisition, and extinction—presented in succession. During habituation, all pictures were presented

Correspondence should be sent to Tiago V. Maia, Department of Psychiatry, Columbia University, 1051 Riverside Drive, Unit 74, New York, NY 10032. E-mail: tmaia@columbia.edu

without reinforcement. During acquisition, one of the pictures from each race (the so-called CS+) was followed by an electric shock, whereas the other (the CS−) was not. During extinction, all pictures were again presented without reinforcement. Conditioning for a given race was obtained by subtracting participants' skin conductance responses (SCRs) to the CS− from their SCRs to the CS+ for that race. Using this difference instead of the SCR to the CS+ controls for possible differences in prior conditioning to each race and for potential sensitization effects (Ohman & Mineka, 2001; but see Lovibond, Siddle, & Bond, 1993). The main finding from the study was that during both acquisition and extinction, participants exhibited more conditioning to members of the other race than to members of their own race.[1] Henceforward, I will refer to this finding as the *superior outgroup conditioning effect*.

Olsson et al. (2005), as well as others (Ohman, 2005), interpreted this finding by appealing to the idea of ''preparedness,'' which proposes that animals (including humans) are genetically programmed to more easily learn to fear stimuli that were dangerous in evolutionary history (Ohman & Mineka, 2001; Seligman, 1971). A substantial body of research has shown that when certain stimuli that were dangerous in evolutionary history (e.g., snakes or crocodiles) are used as the CS in a fear conditioning paradigm, conditioning is stronger and/or more persistent than when a neutral CS (e.g., flowers) is used (Ohman & Mineka, 2001). Similar results have been found for fear conditioning to angry versus happy or neutral faces (Ohman & Dimberg, 1978), and again this has been interpreted in evolutionary terms, as angry faces were presumably more likely to signal danger throughout evolutionary history (Ohman & Dimberg, 1978; Ohman & Mineka, 2001; but see Bond & Siddle, 1996). The findings of Olsson et al. seem to extend these results to a new class of stimulus: social group, as defined by race.

The interpretation of the findings of Olsson et al. (2005) as resulting from evolutionarily determined genetic biases is not without problems, though. As Olsson et al. themselves note, the timing and pattern of differentiation of human groups into what are commonly called human ''races'' (a concept of questionable biological validity; Templeton, 1998) make it unlikely that humans could have evolved mechanisms specifically to learn to fear different races. Human populations differentiated into different races relatively recently in evolutionary history, and more importantly, different groups evolved different characteristics because they were relatively isolated from each other. Being genetically prepared to learn to fear individuals from different races is therefore unlikely to have provided any selective advantage. To get around this problem, Olsson et al. suggest that humans may have evolved instead a more general preparedness to learn to fear ''others who were dissimilar to them or who otherwise appeared not to belong to their social group'' (p. 787). This raises additional difficulties, though. The definition of a social ingroup (and, by exclusion, of a social outgroup) depends on experience and familiarity: Those in my social ingroup are those around me and possibly others similar to them. However, if experience and familiarity are crucial to define the social ingroup, it is unclear that genetic biases play any explanatory role in addressing the findings of Olsson et al.: General-purpose learning mechanisms that are part of the

standard armamentarium of conditioning theories are sufficient to explain higher conditioning to less familiar CSs.

Consider how the genetic biases would work if they depended on learning to define the social groups. There would have to be a learning mechanism that formed a category of ''those around me'' and that could tell the similarity between an individual and that category (or, equivalently, the probability of that individual belonging to the category). The amount of learning about an individual should then be inversely proportional to his or her similarity to that category. In other words, there would be the need for (a) categorization or evaluation of stimulus similarity, and (b) more learning for unfamiliar than for familiar stimuli. Now, both of these mechanisms are part of standard learning theory. Stimulus generalization in classical conditioning was first observed by Pavlov (1927) and has been incorporated in theories of conditioning (McLaren & Mackintosh, 2002). In fact, extending theories of conditioning to account for stimulus generalization is often straightforward, requiring only the use of distributed representations (Hinton, McClelland, & Rumelhart, 1986) for the CSs. There is also an abundance of evidence in the conditioning literature for more learning for unfamiliar CSs. This is illustrated by the phenomenon of *latent inhibition*: the finding that prior exposure to a CS without an accompanying US results in delayed learning when that CS is eventually paired with a US (Lubow, 1973; Lubow & Moore, 1959). The idea that there is more learning for unfamiliar than for familiar CSs has also been incorporated in many theories of conditioning that emphasize the role of attention in determining the associability of stimuli (Dayan, Kakade, & Montague, 2000; Kruschke, 2001; Mackintosh, 1975; Pearce & Bouton, 2001; Pearce & Hall, 1980). Finally, there is also direct evidence for stimulus generalization in latent inhibition (Siegel, 1969), showing that these two mechanisms can be combined. It is therefore unclear that socially or racially specific genetic biases play any explanatory role over and above that provided by the mechanisms in standard learning theories.

There is a further difficulty with interpreting Olsson et al.'s (2005) findings as evidence for prepared social learning. As proponents of preparedness in other domains have emphasized, to conclusively demonstrate prepared learning it is not sufficient to show superior conditioning to the alleged prepared stimulus (CS_1) as compared to a nonprepared stimulus (CS_2) in one context, such as fear conditioning; it is also necessary to show that in a different context, such as appetitive conditioning, CS_1 does *not* show superior conditioning (LoLordo, 1979; Ohman & Mineka, 2001). Otherwise, CS_1 may simply be more salient than CS_2, and it has been known for decades that more salient CSs lead to more learning—a phenomenon that is demonstrated, for example, in overshadowing (Pavlov, 1927) and that is embodied in virtually all theories of conditioning (e.g., Dayan et al., 2000; Kruschke, 2001; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). Unfortunately, the study of Olsson et al. did not address the possibility that, for Black participants, pictures of White individuals may have been more salient than pictures of Black individuals, whereas the opposite may have occurred for White participants. As we will see below, standard learning theory predicts that that should actually have been the case.

2. A latent inhibition account

We saw above that, as extensively explored in several theories that emphasize the role of attention in conditioning, the salience of a stimulus depends not only on the stimulus' characteristics but also on previous experiences with it and with similar stimuli (Dayan et al., 2000; Kruschke, 2001; Mackintosh, 1975; Pearce & Bouton, 2001; Pearce & Hall, 1980). In particular, latent inhibition demonstrates that there is more learning for unfamiliar than for familiar CSs (Lubow, 1973; Lubow & Moore, 1959; Siegel, 1969). Notably, latent inhibition has been demonstrated in humans in conditions very similar to those used in the experiment of Olsson et al. (2005): using aversive USs (including shocks) in a differential fear conditioning paradigm, and with SCRs as the measure of conditioning (Vaitl & Lipp, 1997). This raises the possibility that latent inhibition may have played a role in the experiment of Olsson et al. This section shows that latent inhibition is actually sufficient to account for the findings of Olsson et al.

The participants in Olsson et al.'s (2005) experiment reported significantly more previous exposure to members of their own race than to members of the other race. This suggests that there was significantly more latent inhibition for members of the participants' own race, which in turn is sufficient to explain the finding of stronger conditioning to members of the other race. Importantly, the latent inhibition account would also explain two additional findings in Olsson et al.'s study that the evolutionary explanation does not address. First, Olsson et al. found that contact with members from the other race was the only factor that mediated the superior outgroup conditioning effect.[2] Second, in the graphs in Olsson et al.'s article there was an apparent trend for the superior outgroup conditioning effect to be stronger for White than for Black participants (even though this trend did not reach statistical significance). The latter finding is probably simply a consequence of the former, as Black participants reported more prior exposure to White individuals than White participants did to Black individuals. These two findings are direct predictions of the latent inhibition hypothesis: More contact with members of the outgroup means more latent inhibition for the outgroup, which implies a reduction in the superior outgroup conditioning effect.

The latent inhibition account seems a more parsimonious explanation for the findings of Olsson et al. (2005) for three reasons. First, the latent inhibition account fully explains the superior outgroup conditioning effect. In contrast, as discussed in the introduction, the evolutionary account fails to provide a satisfactory explanation even for that finding; at the very least, it would be necessary to articulate better the interactions between the alleged genetic biases and the mechanisms that would learn about social groups. Second, the latent inhibition account explains additional findings in Olsson et al.'s study that the evolutionary explanation does not address. Third, the latent inhibition account relies on standard learning theory phenomena that have been repeatedly demonstrated over decades of research; in contrast, the evolutionary account suggests a new mechanism—a socially specific genetic bias—for which there is no independent evidence.

While the arguments above establish latent inhibition as a more compelling explanation for the findings of Olsson et al. (2005) than the evolutionary account, a verbal account of

how latent inhibition produces these findings is incomplete. Virtually all theories that explore the role of attention in conditioning explain latent inhibition as a decrease in the associability of preexposed CSs, where the associability of a CS corresponds to the *learning rate* for that CS (Dayan et al., 2000; Mackintosh, 1975; Pearce & Bouton, 2001; Pearce & Hall, 1980). This explains the superior outgroup conditioning effect *during the acquisition phase* in Olsson et al.'s experiment: More prior exposure to the ingroup results in a smaller learning rate for the ingroup, which implies more learning for the outgroup. However, a verbal (i.e., nonquantitative) account cannot fully predict what should happen during extinction. Clearly, at the end of acquisition, conditioning should be greater for the outgroup. However, the learning rates during extinction should still be larger for the outgroup; thus, *unlearning* during extinction should be faster for the outgroup, which could conceivably cancel out the superior outgroup conditioning effect during extinction. Even if this were the case, it would be unlikely to have made a difference in Olsson et al.'s results, given the way they analyzed the data: They averaged the SCRs to all extinction trials except the first one, and it seems unlikely that in a single trial of unlearning the effect of the six acquisition trials would have been totally canceled out. Nevertheless, only a formal, quantitative model can definitively show whether the latent inhibition account can in fact explain the detailed pattern of results of Olsson et al.

3. Conditioning and statistical inference

Much leverage has been gained in understanding conditioning by assuming that animals perform normative statistical inference regarding the relationship between the CSs and the US (e.g., Courville, Daw, & Touretzky, 2006; Dayan et al., 2000; Kruschke, 2008). Under certain reasonable assumptions about the environment (detailed below and in Dayan & Kakade, 2001; Dayan et al., 2000), this statistical inference problem can be solved optimally using the Kalman filter (Brown & Hwang, 1997; Kalman, 1960). Dayan and colleagues have used the Kalman filter to successfully model many conditioning phenomena that standard theories of conditioning have difficulty explaining (Dayan, 1994; Dayan & Kakade, 2001; Dayan et al., 2000). The Kalman filter is therefore well established as an excellent model of conditioning. It also has the advantage of being independently motivated by statistical theory. Unlike other models of conditioning, the Kalman filter model was *not* developed with the aim of fitting empirical findings; instead, it was fully derived on normative statistical grounds. This is important for our purposes, as it might not be too difficult to develop *some* model that would fit the findings of Olsson et al. (2005). By using the Kalman filter, I am able to show that the findings of Olsson et al. are exactly what would be expected if *participants are performing normative statistical inference* regarding the relationship between the CSs and the US in the experiment (given the statistics of their prior exposure to each race). In fact, normative statistical inference explains not only the superior outgroup conditioning effect (during both acquisition and extinction) but also its moderation by outgroup contact and the fact that that effect is stronger for White than for Black participants.

4. The Kalman filter model of conditioning

4.1. Stimulus representation

The CSs present on trial t can be represented by a vector $\mathbf{u}_t$.[3] In the standard Kalman filter model of conditioning, as in most other models of conditioning, each element of the vector $\mathbf{u}_t$ corresponds to a (potential) CS. This type of representation is sufficient to capture many conditioning findings. However, by using a different number to represent each CS, these representations fail to capture the similarity structure between different CSs and are thus unable to exhibit stimulus generalization. Generalization is crucial for the latent inhibition account of the findings of Olsson et al. (2005), though, because participants are unlikely to have had prior exposure to the specific faces used in the experiment. A natural way of obtaining generalization based on similarity is to use a distributed, feature-based representation (Hinton et al., 1986; McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986) to represent CSs, an approach that has previously been used to account for several results concerning stimulus generalization in conditioning (McLaren & Mackintosh, 2002). Accordingly, I represent CSs as a distributed pattern of activation over the vector $\mathbf{u}_t$. Individual CSs—which correspond to the pictures used during the experiment of Olsson et al. (2005) or to individuals that the participants have met in their life prior to the experiment (see Section 5.1)—are generated probabilistically according to the distributions and algorithm in Section S1 of the Supplemental Materials, which are available online at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html. As is standard, I represent the US on trial t as a scalar, y_t. For simplicity, I use binary values for y_t, with 0 and 1 representing the absence and presence of an aversive US, respectively.

4.2. The model

The Kalman filter model assumes that at each time t there is some true set of weights $\mathbf{x}_t$ that reflects the association of $\mathbf{u}_t$ with y_t, according to the following linear equation:

$$y_t = \mathbf{u}_t^T \mathbf{x}_t + v_t, \qquad v_t \sim N(0, R), \tag{1}$$

where v_t represents zero-mean Gaussian noise.

Relations between CS features and the US can change, so it is important to allow $\mathbf{x}_t$ to change with time. In the absence of new observations, the animal's best prediction about the relation between CS features and the US at time $t + 1$ is that it is the same as it was at time t (i.e., $\mathbf{x}_{t+1} = \mathbf{x}_t$). However, there is also some probability that the weights have drifted away from $\mathbf{x}_t$. These ideas are captured by the following dynamic equation (Dayan et al., 2000):

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q}), \tag{2}$$

where $\mathbf{w}_t$ is a vector of Gaussian noise with mean $\mathbf{0}$ and covariance matrix $\mathbf{Q}$. I made $\mathbf{Q}$ equal to $c_1\mathbf{I}$, where c_1 is a constant and $\mathbf{I}$ is the identity matrix, to ensure that *there are no ''innate'' biases suggesting that some weights change more over time than others or that changes in different weights are correlated.*

To finish specifying the system, it is necessary to define a probability distribution over the initial weights $\mathbf{x}_0$. In the Kalman filter, this distribution is Gaussian:

$$\mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{\Sigma}_0). \tag{3}$$

I made the mean $\mathbf{0}$ to ensure that *there are no ''innate'' biases suggesting that certain features are more likely to be associated with the US*. I made $\mathbf{\Sigma}_0$ equal to $c_2\mathbf{I}$, where c_2 is a constant and $\mathbf{I}$ is the identity matrix, to ensure that *there are no ''innate'' biases concerning the uncertainty (i.e., the salience) of the different features*.[4]

4.3. Inferring the relation between the conditioned stimuli and the unconditioned stimulus

The organism's inference problem is to infer the probability distribution over $\mathbf{x}_t$, given all previously seen pairings $(\mathbf{u}_1, y_1)$, $(\mathbf{u}_2, y_2)$, $\ldots$, $(\mathbf{u}_t, y_t)$. Given that the probability distribution for $\mathbf{x}_0$ is Gaussian and that all equations are linear, $\mathbf{x}_t$ will have a Gaussian distribution for all t. The inference problem therefore reduces to finding the mean $\hat{\mathbf{x}}_t$ and covariance matrix $\mathbf{\Sigma}_t$ of that Gaussian distribution. Kalman (1960) devised recursive equations that determine $\hat{\mathbf{x}}_t$ and $\mathbf{\Sigma}_t$ as a function of $\hat{\mathbf{x}}_{t-1}$, $\mathbf{\Sigma}_{t-1}$, and observation $(\mathbf{u}_t, y_t)$. These equations are presented in Section S2 of the Supplemental Materials. Repeated application of the Kalman filter update equations from times 1 through t calculates the posterior distribution over $\mathbf{x}_t$ [i.e., $N(\mathbf{x}_t, \mathbf{\Sigma}_t)$], given the prior distribution over $\mathbf{x}_0$ [i.e., $N(\mathbf{x}_0, \mathbf{\Sigma}_0)$] and observations $(\mathbf{u}_1, y_1)$, $(\mathbf{u}_2, y_2)$, $\ldots$, $(\mathbf{u}_t, y_t)$.

The most important aspect of the Kalman filter for our purposes is that the associabilities (i.e., learning rates) for each CS feature are a function of *uncertainty*. The more uncertainty there is about a particular weight (i.e., about the relation between a particular CS feature and the US), the more learning there is for that weight when that CS feature is present. This is intuitive: If we are very certain about the relation between a particular CS feature and the US, a new observation will have a smaller effect than if we are very uncertain about that relation. This occurs in any Bayesian approach: If the prior is strongly peaked (i.e., if the uncertainty is low), new observations count less than if the prior is flatter. In the Kalman filter model, the uncertainty for particular CS features in turn depends on how often one has seen those features: Uncertainty about particular CS features diminishes as one encounters those features. The key point, then, is that the less often one has seen particular CS features, the higher the uncertainty—and therefore the higher the learning rate—for those features. In other words, familiarity leads to a lower learning rate, and unfamiliarity leads to a higher learning rate.[5] (For the mathematical details explaining why learning rates depend on uncertainties, and why uncertainties in turn depend on prior exposure, please refer to Section S2 of the Supplemental Materials.)

4.4. Implications for latent inhibition and for the findings of Olsson et al. (2005)

We can now understand the statistical basis for latent inhibition and for the findings of Olsson et al. (2005). Consider first the standard latent inhibition finding (Lubow, 1973;

Lubow & Moore, 1959). To account for this finding, we do not even need a distributed representation for the CSs. Suppose that each (potential) CS is represented by a distinct element in vector $\mathbf{u}_t$. When the animal is preexposed to a given CS, the uncertainty about the weight associated with that CS decreases; therefore, when that CS is eventually paired with a US, learning is slower than for a nonpreexposed CS. Next, consider the finding of stimulus generalization in latent inhibition (Siegel, 1969). Accounting for this finding requires a distributed representation for the CSs, but the explanation is exactly the same: Preexposure to a given CS results in a reduction in the uncertainty associated with the features of that CS, which in turn leads to slower learning not only for that CS but also for other stimuli with similar features. Furthermore, the more similar the test CSs are to the preexposed CS, the less learning there should be, as is found empirically (Siegel, 1969).

The explanation for the results of Olsson et al. (2005) is similar. Participants have less prior exposure to the outgroup, so their uncertainty about the weights associated with the features that are characteristic of the outgroup is higher. This, in turn, means that learning is higher when participants are shown a picture of an individual from the outgroup, which results in the superior outgroup conditioning effect. The next section shows quantitatively that this statistical explanation accounts in detail for the findings of Olsson et al.

5. Simulation 1: The main findings of Olsson et al. (2005)

I simulated the experiment of Olsson et al. (2005) using the same experimental parameters that they used (including the same number of participants, same number of trials in each phase, etc.). Crucially, in the simulations Black and White participants were represented by *exactly the same* statistically normative Kalman filter model. In other words, there were no ''innate biases'' built into the simulations; the only difference between Black and White participants was the amount of exposure they had to each race prior to the experiment.

For each participant, I generated probabilistically an individual ''life history'' of ($\mathbf{u}_t$, y_t) pairs (see Section 5.1). The Kalman filter was applied to this history to simulate the participant's experience before the experiment. Each participant therefore entered the experiment with different estimates of the probability distribution over the weights that relate the CS features to the USs (i.e., different estimates $\hat{\mathbf{x}}_s$ and $\mathbf{\Sigma}_s$, where s represents the time at which the experiment starts).[6] Importantly, such estimates depended entirely on the participant's history and not on any innate biases. The participant (i.e., the Kalman filter model with initial estimates $\hat{\mathbf{x}}_s$ and $\mathbf{\Sigma}_s$) was then exposed to the experimental contingencies (see Section 5.2).

5.1. Simulation of participants' life history

As mentioned above, I generated an individual ''life history'' of ($\mathbf{u}_t$, y_t) pairs for each participant. The statistics of these life histories were consistent with the self-reported data in Olsson et al. (2005): (a) Black participants had more exposure to Black than to White individuals; (b) White participants had more exposure to White than to Black individuals; and (c) these differences were more pronounced for White than for Black participants

(i.e., White participants had less exposure to Black individuals than Black participants had to White individuals). For simplicity, all participants had the same number N of $(\mathbf{u}_t, y_t)$ pairs in their life histories. In accordance with the way Olsson et al. (2005) coded their data, Black and White were treated as discrete, mutually exclusive categories. As no other races were involved in the experiment, these two categories were also exhaustive.

Let Z_t be the random variables that represent whether at time t ($t = 1...N$) a given participant is meeting a Black or a White individual (represented by 1 and 0, respectively). I modeled Z_t as coming from a Bernoulli distribution and considered that $Z_1, Z_2, \ldots, Z_N$ are independent and identically distributed (so, they form Bernoulli trials). The parameter for Z_t depends on the race of the participant; this captures the difference in experience between Black and White participants. If we let r represent the participant's race, we are interested in the distributions $p(Z_t \mid r)$. Both $p(Z_t \mid r = \mathrm{B})$ and $p(Z_t \mid r = \mathrm{W})$ are Bernoulli distributions, but with different parameters. I will represent these parameters by $p_{\mathrm{B|B}}$ (the probability of the participant meeting a Black individual given that the participant is Black) and $p_{\mathrm{B|W}}$ (the probability of the participant meeting a Black individual given that the participant is White), respectively.

The exact values that we should give to $p_{\mathrm{B|B}}$ and $p_{\mathrm{B|W}}$ are open; however, to be consistent with the self-reported data in Olsson et al. (2005), these values should obey the following constraints:

1. Given that Black participants had more exposure to Black than to White individuals, $p_{\mathrm{B|B}} > .5$.
2. Given that White participants had more exposure to White than to Black individuals, $p_{\mathrm{B|W}} < .5$.
3. Given that Black participants had more exposure to White individuals than White participants had to Black individuals, $p(Z_t = 0 \mid r = \mathrm{B}) > p(Z_t = 1 \mid r = \mathrm{W})$, which can also be written as $1 - p_{\mathrm{B|B}} > p_{\mathrm{B|W}}$.

Given that my emphasis is on demonstrating that latent inhibition is sufficient to account for the findings of Olsson et al. (2005), I made aversive experiences equiprobable for both races, regardless of the race of the participant. In other words, the probability p_{US} that y_t was 1 in the participant's life history was fixed, regardless of the race of the CS ($\mathbf{u}_t$) or the race of the participant. Note, though, that this does not affect our results: Any differences in conditioning prior to the experiment would be canceled out by the experiment's differential conditioning design.

Section S3 of the Supplemental Materials presents the algorithm used to generate participants' life histories, as well as the specific values that I used for the parameters introduced in this section (N, $p_{\mathrm{B|B}}$, $p_{\mathrm{B|W}}$, and p_{US}).

5.2. Simulation of the experiment

I simulated 37 Black and 36 White participants to match the number of participants in the experiment of Olsson et al. (2005). The simulated experiment had exactly the same

parameters as the experiment of Olsson et al. It consisted of three phases: an initial habituation phase, in which participants saw four presentations of each CS without the US; an acquisition phase, in which participants saw six presentations of each CS, with every presentation of the CS+s accompanied by the US; and an extinction phase, in which participants saw six presentations of each CS without the US. As in the experiment of Olsson et al., the order of presentation of the CSs within each phase was randomized across participants. The four stimuli for the experiment (two pictures of Black individuals and two pictures of White individuals) were generated according to the algorithm in Section S1 of the Supplemental Materials. As in the experiment of Olsson et al., the same stimuli were used for every participant, but the role that each stimulus played (i.e., whether it was a CS+ or a CS−) was randomly selected for each participant.

5.3. Simulation of the skin conductance responses

In the Kalman filter model, the relation between the CSs or CS features and the US is linear. The electrodermal system, however, does not respond linearly (Boucsein, 1992). To simulate the nonlinearity in the electrodermal response system, I simulate SCRs by applying a logistic transformation to the predictions of the Kalman filter (see Section S4 of the Supplemental Materials).

I analyzed the simulated SCR data in the same way that Olsson et al. (2005) analyzed their SCR data. Specifically, the mean SCRs for each CS during each phase of the experiment were calculated as follows: During habituation, the mean SCR was calculated as the mean of the four habituation presentations of that CS. During acquisition, the mean SCR was calculated as the mean of all acquisition trials for that CS except the first one and including the first extinction trial. During extinction, the mean SCR was calculated as the mean of all extinction trials for that CS except the first one.

5.4. Results

5.4.1. The superior outgroup conditioning effect

We are interested in whether the simulation results, like the results of Olsson et al. (2005), exhibit the superior outgroup conditioning effect during both acquisition and extinction. Fig. 1 shows that conditioning is indeed higher for the outgroup than the ingroup during both acquisition and extinction. Paired t-tests confirm that these differences are significant during both acquisition, $t(72) = 13.12$, $p < .01$ (two-tailed), and extinction, $t(72) = 9.80$, $p < .01$ (two-tailed).

5.4.2. The superior outgroup conditioning effect for Black and White participants

We are also interested in whether the simulation results, like the results of Olsson et al. (2005), exhibit the superior outgroup conditioning effect during both acquisition and extinction for both Black and White participants. Fig. 2 shows that conditioning is indeed higher for the outgroup than for the ingroup during both acquisition and extinction, for both Black and White participants. Paired t-tests confirm that these differences are significant: For

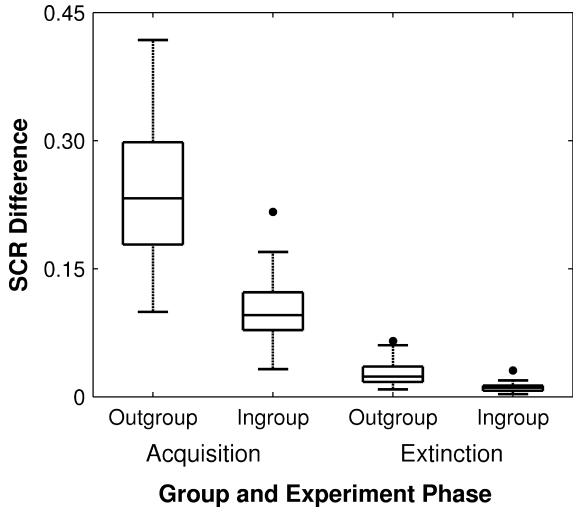

Fig. 1. Box plots showing conditioning to the outgroup and the ingroup during acquisition and extinction. In this figure, as in all subsequent box plots in this article, the boxes have lines at the lower quartile, median, and upper quartile; the maximum whisker length is 1.5 × interquartile range, and outliers are represented by dots. The SCR difference is the difference between the SCR for the CS+ and the SCR for the CS−.

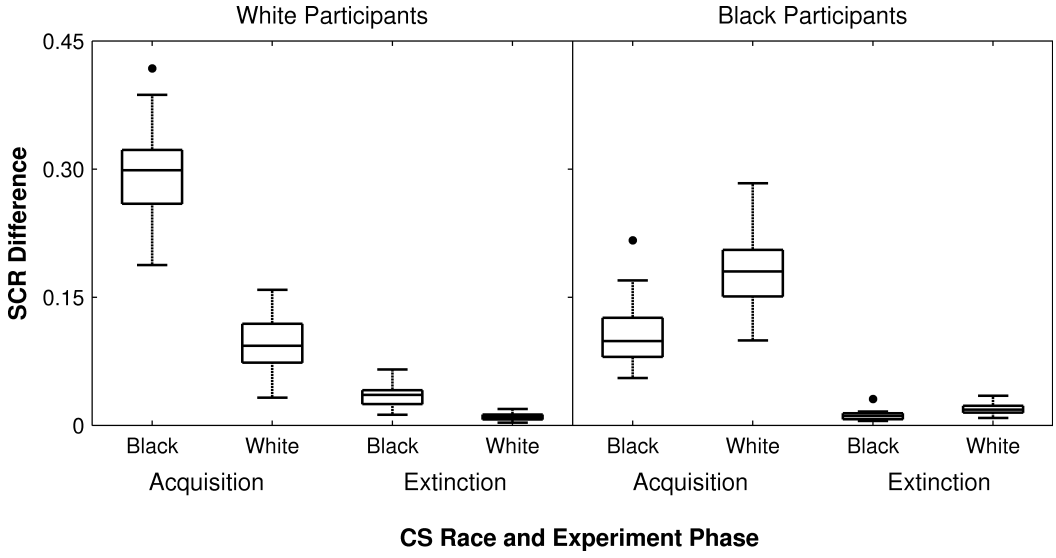

Fig. 2. Box plots showing conditioning to the pictures of Black and White individuals during acquisition and extinction, presented separately by participant race.

Black participants, $t(36) = 6.94$, $p < .01$, during acquisition, and $t(36) = 5.27$, $p < .01$, during extinction; for White participants, $t(35) = 19.17$, $p < .01$, during acquisition, and $t(35) = 11.03$, $p < .01$, during extinction (all two-tailed).

5.4.3. The strength of the superior outgroup conditioning effect for Black and White participants

Finally, we are interested in whether in the simulation results, like in the results of Olsson et al. (2005), the superior outgroup conditioning effect is stronger for White than for Black participants, during both acquisition and extinction. This is already apparent in Fig. 2, which suggests that the difference in conditioning to the outgroup and the ingroup is more pronounced for White than for Black participants, during both acquisition and extinction. Fig. 3 shows more directly that this difference (i.e., the superior outgroup conditioning effect) is indeed more pronounced for White than for Black participants during both acquisition and extinction. Between-subjects *t*-tests confirm that the strength of the superior outgroup conditioning effect is significantly different for Black and White participants during both acquisition, $t(71) = 8.11$, $p < .01$ (two-tailed), and extinction, $t(71) = 6.16$, $p < .01$ (two-tailed).

5.4.4. Robustness of findings

Additional simulations demonstrate that all of these findings still hold when the parameters governing the generation of life histories and CSs are chosen randomly from the range of plausible values (see Section S5 of the Supplemental Materials).

5.4.5. The key to understanding the findings: Learning rates

The findings in Figs. 1–3 are a direct result of the statistics of participants' experiences. The fact that participants have less prior exposure to individuals from the outgroup means that at the beginning of the experiment they have higher learning rates for the outgroup. This is illustrated in Fig. 4. That figure also illustrates why the superior outgroup conditioning effect is stronger for White than for Black participants: Because Black participants had more prior exposure to White individuals than vice-versa, the learning rates for pictures of Black and White individuals are more similar for Black participants than they are for White participants. Finally, note in the figure that the learning rates decrease throughout the

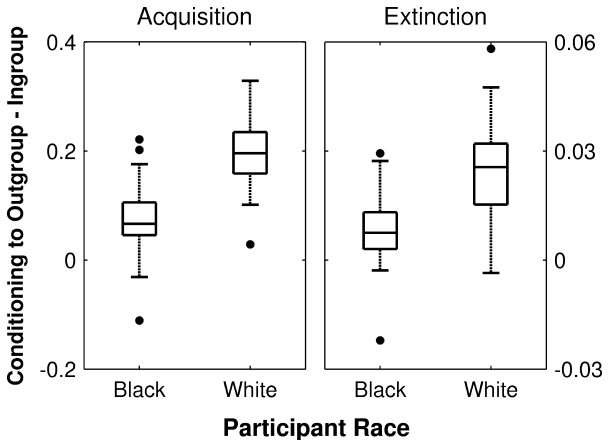

Fig. 3. Box plots of the difference in conditioning to the outgroup and the ingroup during acquisition and extinction, presented separately for Black and White participants.

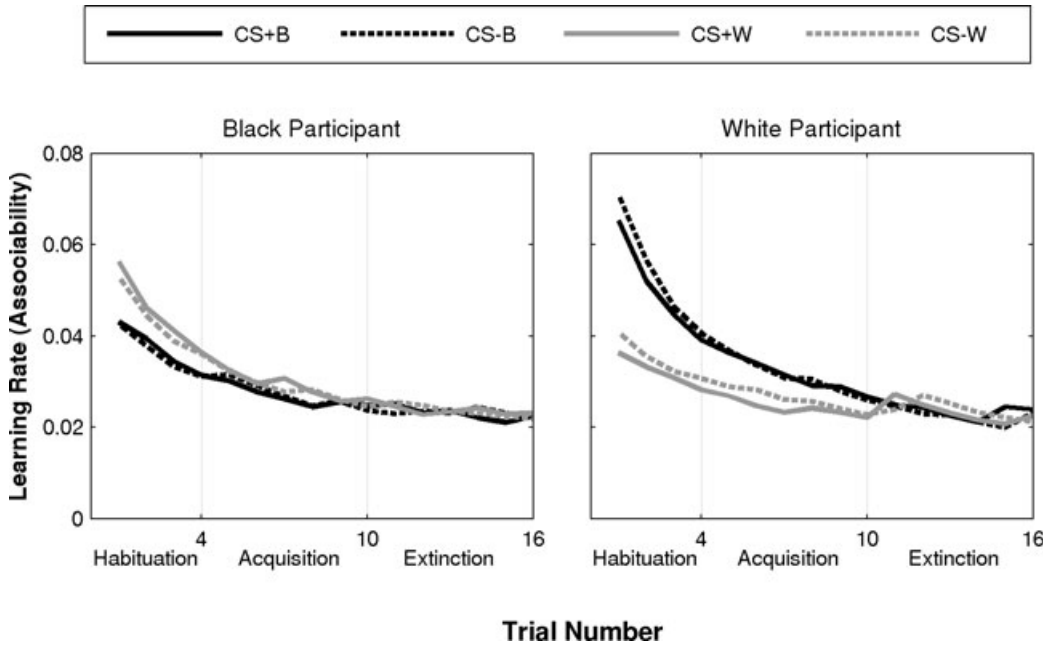

Fig. 4. Associability (learning rate) for the different types of stimuli throughout the experiment, for an example Black participant (left panel) and an example White participant (right panel). Each line represents the associability for a particular type of stimulus (CS+B: picture of Black individual that served as a CS+; CS−B: picture of Black individual that served as a CS−; CS+W: picture of White individual that served as a CS+; CS−W: picture of White individual that served as a CS−). Both the Black and the White participants have higher associabilities for the outgroup (the different race) at the beginning of the experiment; however, this effect is significantly more pronounced for the White participant. Note also that the associability for the CS+ and the CS− of a given race are very similar throughout the experiment. This is because in the Kalman filter model of conditioning, associabilities depend only on the number of exposures to a given CS and not on whether that CS was associated with a US. For details on how the associability of a CS was calculated, see Section S6 of the Supplemental Materials.

experiment. This is due to repeated exposure to the stimuli presented in the experiment, which reduces the uncertainty for those stimuli. Importantly, the Kalman filter formulas make larger learning rates decrease more sharply, so the learning rates for the outgroup and the ingroup become more similar as the experiment progresses. This means that whereas acquisition will occur significantly faster for the outgroup, extinction may be only slightly faster for the outgroup, if at all.

Recall that a verbal account based on latent inhibition had difficulty predicting whether the superior outgroup conditioning effect would wash away entirely during extinction, given that learning during extinction (i.e., *unlearning* of the fear) should be faster for the outgroup (see Section 2). The Kalman filter model shows, on normative statistical grounds, that whereas during extinction the learning rate for the outgroup should indeed, on average, be higher than the learning rate for the ingroup, the difference in learning rates between the outgroup and the ingroup should be smaller during extinction than during acquisition. Hence, one would expect to see the superior outgroup conditioning effect also during extinction.

6. Simulation 2: The effects of contact

In an additional simulation, I addressed specifically the effect of contact with members from each race on the differences in conditioning to each race. In this simulation, the percentage of each participant p's prior exposure to White individuals, $p_W(p)$, was generated randomly from a uniform distribution between 0 and 100; the percentage of p's exposure to Black individuals was $100 - p_W(p)$. (Recall that in Simulation 1 the only difference between Black and White participants was in their prior exposure to each race; in the current simulation, the amount of prior exposure to each race was generated randomly for each participant, so there was no distinction between Black and White participants.) Participant p's life history was generated as in Simulation 1, but using $p_W(p)/100$ as the parameter for the Bernoulli variables Z_t that represent whether at time t ($t = 1...N$), p meets a Black or a White individual (see Section 5.1). All other aspects of the simulation were the same as in Simulation 1. There were 50 participants in the simulation.

I measured how much stronger conditioning was to pictures of Black individuals than to pictures of White individuals by subtracting the conditioning to pictures of White individuals from the conditioning to pictures of Black individuals. As expected, I found a strong and significant positive correlation between the amount of prior exposure to White individuals and the extent to which conditioning was stronger for Black than for White individuals, during both acquisition, $r(48) = .89$, $p < .01$, and extinction, $r(48) = .78$, $p < .01$ (see Fig. 5). Note that the higher the proportion of exposure to White individuals, the smaller the learning rate for White individuals relative to the learning rate for Black individuals. This, in turn, results in a larger difference in conditioning to the pictures of Black individuals minus conditioning to the pictures of White individuals. This simulation illustrates Olsson et al.'s (2005) finding that the superior outgroup conditioning effect is mediated by the amount of contact with each race.

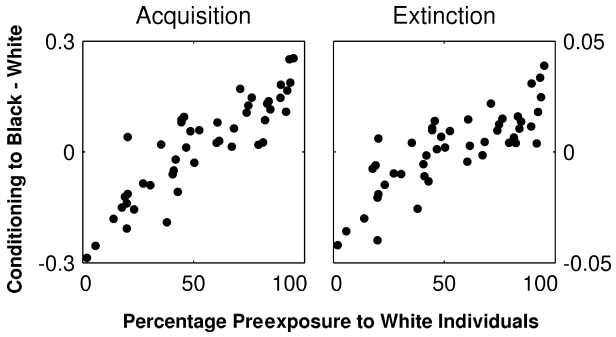

Fig. 5. Scatter plot showing conditioning to pictures of Black individuals minus conditioning to pictures of White individuals as a function of the amount of preexposure to White individuals. Each point in the graph corresponds to a participant.

7. Discussion

7.1. Alternative explanations

In addition to the preparedness and statistical/latent inhibition explanations for the findings of Olsson et al. (2005), which I have emphasized thus far, we also need to consider two other alternative explanations. The first is that those findings may be the product of social and cultural learning (Ohman, 2005). Social/cultural learning in this context is generally thought of in terms of learning of explicit or implicit attitudes or stereotypes (Cunningham, Preacher, & Banaji, 2001; Greenwald & Banaji, 1995). However, Olsson et al. collected several measures of implicit and explicit stereotyping and found that those did not correlate with the superior outgroup conditioning effect. Furthermore, such social and cultural learning would presumably apply equally to the CS+ and CS− from each race, so when the SCR for the CS− is subtracted from the SCR for the CS+, any such effects should cancel out. Overall, then, this does not seem a likely explanation for the findings of Olsson et al. A second, related explanation is that these findings may be due to different prior conditioning to members of the ingroup and the outgroup. However, again, by subtracting conditioning to the CS− from conditioning to the CS+ for each race, the experimental paradigm used by Olsson et al. controls for that possibility. Furthermore, if there had been higher preconditioning to the outgroup than to the ingroup, such effect should be apparent during the habituation phase. However, SCRs were actually slightly larger for the ingroup than for the outgroup during habituation, which does not support the idea that there was more conditioning to the outgroup at the beginning of the experiment.

7.2. The superior outgroup conditioning effect in participants without outgroup dating experience

In addition to the findings discussed above, Olsson et al. (2005) also presented an analysis restricted to participants without outgroup dating experience. As expected, both Black and White participants showed higher conditioning to the outgroup, but unlike in the main analysis, the effect did not seem stronger for White participants. To account for these differences between analyses, the genetic preparedness hypothesis would have to postulate another, independent mechanism mediating the effects of outgroup dating. In contrast, the latent inhibition hypothesis offers a natural explanation for these findings, as we will now see.

Unfortunately, Olsson et al. did not present the results of the measures of outgroup contact for participants without outgroup dating experience. Nevertheless, outgroup dating correlated strongly with all other measures of outgroup contact, so removing participants with outgroup dating experience from the analysis skewed the sample toward less contact with the outgroup. Importantly, many more Black than White participants reported interracial dating (51% vs. 28%, respectively), so such skewing affected the sample of Black participants much more than it affected the sample of White participants. As a result,

differences in outgroup contact between Black and White participants were likely to have been greatly reduced, if they persisted at all, when participants with outgroup dating experience were removed from the sample. Under the latent inhibition account, the differences in strength of the superior outgroup conditioning effect for Black and White participants should therefore have been greatly reduced, or even abolished, when participants with outgroup dating experience were removed from the analysis. This is exactly what Olsson et al. found.

7.3. Predictions and tests of the model

The account developed in this article can be tested empirically in a variety of ways. This section suggests three empirical tests that would differentiate this account from the preparedness hypothesis.

7.3.1. Pharmacological manipulations that affect latent inhibition

Drugs that increase dopamine, such as amphetamines, disrupt latent inhibition, whereas dopamine receptor antagonists, such as haloperidol and other antipsychotic drugs, increase latent inhibition (for review, see, e.g., Lubow, 2005; Moser, Hitchcock, Lister, & Moran, 2000; Young, Moran, & Joseph, 2005). Although most studies have administered the drugs prior to both preexposure and conditioning, drug administration prior to conditioning alone seems sufficient to affect latent inhibition (Joseph, Peters, & Gray, 1993; Joseph et al., 2000; Moser et al., 2000; Peters & Joseph, 1993; Young et al., 2005). The latent inhibition account could therefore be tested directly by repeating the experiment of Olsson et al. (2005) with participants given a drug that disrupts latent inhibition. If the drug successfully abolished latent inhibition, the latent inhibition hypothesis predicts that the superior outgroup conditioning effect would also be abolished. The genetic preparedness hypothesis, in contrast, predicts no effect of these pharmacological manipulations on the superior outgroup conditioning effect.

7.3.2. Manipulations of habituation

As illustrated in Fig. 4, the Kalman filter model predicts that during habituation the associability decreases more sharply for members of the outgroup than for members of the ingroup. As a result, habituation trials tend to make learning for the outgroup and the ingroup more similar. Increasing the number of habituation trials should therefore lead to a reduction in the superior outgroup conditioning effect. In other words, increasing the number of habituation trials should lead to a greater decrease in conditioning to the outgroup than in conditioning to the ingroup. The genetic preparedness hypothesis, in contrast, would predict that increasing the number of habituation trials would either affect conditioning to the outgroup *less* than conditioning to the ingroup (because prepared stimuli might be less susceptible to habituation), or, at the very least, that conditioning to the outgroup and the ingroup would be affected equally (because preparedness and habituation would be independent effects). An experiment that varied

the number of habituation trials could be used to test these differential predictions of the two accounts.

7.3.3. Appetitive conditioning

The latent inhibition account predicts that in an experiment similar to that of Olsson et al. (2005), but using an appetitive rather than aversive US, conditioning would also be higher for the outgroup. The genetic preparedness hypothesis, in contrast, would predict either no differences between the groups or higher conditioning for the ingroup.

7.4. Related work

Bond and Siddle (1996) have offered a latent inhibition account for the finding of stronger fear conditioning to angry versus happy or neutral faces, which had previously been interpreted as evidence for prepared learning (Ohman & Dimberg, 1978). Bond and Siddle (1996) hypothesized that humans learn more easily to fear angry than happy or neutral faces simply because angry faces are less common, and in a study with different facial expressions found evidence in support of that hypothesis.

The statistical account developed in this article is closely related to latent inhibition. However, the statistical account goes beyond a verbal account that just appeals to latent inhibition, in two main ways: (a) It provides a normative justification for the higher learning for unfamiliar stimuli, and (b) it explains why superior conditioning to the outgroup does not entirely disappear during extinction.

8. Conclusions

I have shown that general-purpose learning mechanisms that perform normative statistical inference (Kalman, 1960) and that have been used to account for a wealth of other findings in conditioning (Dayan, 1994; Dayan & Kakade, 2001; Dayan et al., 2000) offer a parsimonious and integrated explanation of all aspects of the findings of Olsson et al. (2005). The explanation based on socially specific genetic biases, in contrast, requires postulating genetic influences for which no independent evidence exists, and it fails to address certain aspects of the findings (e.g., the effect of outgroup contact on the superior outgroup conditioning effect). Olsson et al. suggested that ''[m]illennia of natural selection and a lifetime of social learning may predispose humans to fear those who seem different from them'' (p. 787). The account developed in this article is very different. The optimal statistical solution to the problem of inferring the relation between CSs and USs requires more learning for less familiar stimuli. When, as a group, participants in a classical conditioning experiment are less familiar with people from other races (as in the experiment of Olsson et al.), conditioning will naturally be stronger when the CS is a picture of an individual from another race than when it is a picture of an individual from the participant's own race. This explanation is consistent with a wealth of findings regarding latent inhibition in the animal

and human literatures (Lubow, 1973; Lubow & Moore, 1959; Siegel, 1969; Vaitl & Lipp, 1997).

Notes

1. Olsson et al. (2005) emphasized this finding during extinction, but it was statistically significant during both acquisition and extinction.
2. To assess ingroup and outgroup contact, Olsson et al. (2005) asked participants how many ingroup and outgroup dating partners they had had, as well as how many ingroup and outgroup friends and acquaintances they had. To obtain relative measures of contact with the outgroup, Olsson et al. subtracted contact with the ingroup from contact with the outgroup, for each of the three measures (dating, friends, and acquaintances). Each of the three resulting measures of contact with the outgroup exhibited a negative correlation with the superior outgroup conditioning effect, as predicted by the latent inhibition account. The only correlation that reached statistical significance, however, was the one involving dating. Interestingly, this too can be explained by the latent inhibition account: People often have more contact with their dating partner than with friends or acquaintances, so outgroup dating should have a stronger impact than outgroup friends or acquaintances on latent inhibition for the outgroup.
3. All vectors in this article are column vectors. When a row vector is desired, vector transposition is explicitly used (e.g., $\mathbf{u}_t^T$).
4. All simulations in this article used the following parameter values: $c_1 = 0.001$, to represent a slowly varying environment; $c_2 = 10{,}000$, to represent a large initial uncertainty about the weights; and $R = 2$, which represents a standard deviation in the estimate of reward of $\sqrt{2}$.
5. Things are more complicated than this simplified explanation may suggest, because in general the covariance matrix for the distribution over $\mathbf{x}_t$ is not diagonal. Thus, it is not only the uncertainty (i.e., variance) for a given weight that is important but also its covariance with other weights. Nevertheless, the intuitive understanding conveyed here is sufficient to understand all the results in this article.
6. There is obviously no claim that real participants have *conscious* estimates of these quantities.

Acknowledgments

I would like to thank Peter Dayan, Nathaniel Daw, and Yael Niv for very helpful discussions regarding the Kalman filter model of conditioning, and Jay McClelland for very helpful discussions regarding this project as a whole. This work was supported in part by a fellowship from the Calouste Gulbenkian Foundation (Portugal).

References

Bond, N. W., & Siddle, D. A. T. (1996). The preparedness account of social phobia: Some data and alternative explanations. In R. M. Rapee (Ed.), *Current controversies in the anxiety disorders* (pp. 291–316). New York: Guilford Press.

Boucsein, W. (1992). *Electrodermal activity*. New York: Plenum Press.

Brown, R. G., & Hwang, P. Y. C. (1997). *Introduction to random signals and applied Kalman filtering: With Matlab exercises and solutions* (3rd ed.). New York: Wiley.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*, 294–300.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170.

Dayan, P. (1994). *Surprising predictions*. Retrieved February 2, 2006, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.8217.

Dayan, P., & Kakade, S. (2001). Explaining away in weight space. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 451–457). Cambridge, MA: MIT Press.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3* (Suppl.), 1218–1223.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 77–109). Cambridge, MA: MIT Press.

Joseph, M. H., Peters, S. L., & Gray, J. A. (1993). Nicotine blocks latent inhibition in rats: Evidence for a critical role of increased functional activity of dopamine in the mesolimbic system at conditioning rather than pre-exposure. *Psychopharmacology*, *110*, 187–192.

Joseph, M. H., Peters, S. L., Moran, P. M., Grigoryan, G. A., Young, A. M., & Gray, J. A. (2000). Modulation of latent inhibition in the rat by altered dopamine transmission in the nucleus accumbens at the time of conditioning. *Neuroscience*, *101*, 921–930.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, *82*, 35–45.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior*, *36*, 210–226.

LoLordo, V. M. (1979). Selective associations. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and memory: A memorial volume to Jerzy Konorski* (pp. 367–398). Hillsdale, NJ: Erlbaum.

Lovibond, P. F., Siddle, D. A., & Bond, N. W. (1993). Resistance to extinction of fear-relevant stimuli: Preparedness or selective sensitization? *Journal of Experimental Psychology: General*, *122*, 449–461.

Lubow, R. E. (1973). Latent inhibition. *Psychological Bulletin*, *79*, 398–407.

Lubow, R. E. (2005). Construct validity of the animal latent inhibition model of selective attention deficits in schizophrenia. *Schizophrenia Bulletin*, *31*, 139–153.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced pre-exposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology*, *52*, 415–419.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models*. Cambridge, MA: MIT Press.

McLaren, I. P., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning and Behavior*, *30*, 177–200.

Moser, P. C., Hitchcock, J. M., Lister, S., & Moran, P. M. (2000). The pharmacology of latent inhibition as an animal model of schizophrenia. *Brain Research. Brain Research Reviews*, *33*, 275–307.

Ohman, A. (2005). Conditioned fear of a face: A prelude to ethnic enmity? *Science*, *309*, 711–713.

Ohman, A., & Dimberg, U. (1978). Facial expressions as conditioned stimuli for electrodermal responses: A case of ''preparedness''? *Journal of Personality and Social Psychology*, *36*, 1251–1258.

Ohman, A., Frederikson, M., Hugdahl, K., & Rimmo, P. A. (1976). The premise of equipotentiality in human classical conditioning: Conditioned electrodermal responses to potentially phobic stimuli. *Journal of Experimental Psychology: General*, *105*, 313–337.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*, 483–522.

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005). The role of social groups in the persistence of learned fear. *Science*, *309*, 785–787.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*, 111–139.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552.

Peters, S. L., & Joseph, M. H. (1993). Haloperidol potentiation of latent inhibition in rats: Evidence for a critical role at conditioning rather than pre-exposure. *Behavioural Pharmacology*, *4*, 183–186.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Seligman, M. (1971). Phobias and preparedness. *Behavior Therapy*, *2*, 307–320.

Siegel, S. (1969). Generalization of latent inhibition. *Journal of Comparative and Physiological Psychology*, *69*, 157–159.

Templeton, A. R. (1998). Human races: A genetic and evolutionary perspective. *American Anthropologist*, *100*, 632–650.

Vaitl, D., & Lipp, O. V. (1997). Latent inhibition and autonomic responses: A psychophysiological approach. *Behavioural Brain Research*, *88*, 85–93.

Young, A. M., Moran, P. M., & Joseph, M. H. (2005). The role of dopamine in conditioning and latent inhibition: What, when, where and how? *Neuroscience and Biobehavioral Reviews*, *29*, 963–976.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplemental Materials.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.