12 Animal Models of Human Attitudes: Integrations Across Behavioral, Cognitive, and Social Neuroscience

Elizabeth A. Phelps and Mahzarin R. Banaji

The development of brain imaging techniques has led to rapid advancement in our understanding of the human brain. Initially, this growth relied on detailed models of neural systems in nonhuman animals to help interpret the function and significance of brain signals observed in humans (Cohen, Noll, & Schneider, 1993). However, as we move from investigating the neural mechanisms of simple perceptual, motor, and cognitive behaviors to complex social interactions, the usefulness of referring to animal models may seem less apparent. This is primarily because human social behavior differs fundamentally from the types of behavior typically observed in laboratory animals. Evidence suggests, however, that even in the study of complex social behavior and culture-specific learning, animal models may provide a useful starting point as we attempt to identify the neural systems underlying these behaviors.

The basic processes of emotional learning and expression are similar across species, although the content and complexity of stimuli may vary widely. By using animal models of emotional learning as a basis for studying complex social responses, we can take advantage of previously identified neural and behavioral mechanisms to inform our understanding of human social interaction. In our experiments, we have benefited from using the mechanisms of classical fear conditioning, which have been investigated across species, to explore affective states in humans that reveal race bias.

Classical conditioning was first described by Pavlov over a century ago, but more recently aversive classical conditioning paradigms, or fear conditioning, have been used to delineate the neural systems of emotional learning (see LeDoux, 2000, for a review). In fear conditioning, a neutral stimulus, such as a tone, is paired with an aversive event, such as a mild shock. After a few pairings, the subject learns that the tone, or conditioned stimulus (CS), predicts the shock, or unconditioned stimulus (UCS), and the tone itself begins to elicit a fear response, called the conditioned response (CR). Studies examining fear conditioning in rats have traced the pathways of emotional learning from stimulus input to response output (Davis, 1997; Kapp et al., 1979;

LeDoux, 1996). These models identified the amygdala as a critical structure in the acquisition, storage, and expression of conditioned fear (LeDoux, 2002; see also Cahill et al., 1999).

Although it is not possible to study neural mechanisms in humans with the same level of specificity as in other animals, both human brain imaging and lesion studies are consistent with animal models. Using functional magnetic resonance imaging (fMRI), amygdala activation was observed during fear conditioning (Buchel et al., 1998; LaBar et al., 1998). Moreover, the strength of that activation is correlated with the CR, as assessed with skin conductance (SCR), an indication of physiological arousal (LaBar et al., 1998). Furthermore, consistent with animal models suggesting the critical involvement of the amygdala in the acquisition and expression of conditioned fear, humans with damage to this region fail to demonstrate a CR as assessed with SCR (Bechara et al., 1995; LaBar et al., 1995).

Human beings, however, can also be asked to report on their experience of emotion. It appears that even though amygdala damage in humans impairs the physiological expression of conditioned fear, these same patients are able to report explicitly on the events that constitute fear conditioning. For example, patient SP, who suffers from bilateral amygdala damage, was given several presentations of a blue square paired with a mild shock to the wrist. Normal control subjects show an increased SCR to the blue square after a few pairings with the shock, an indication of conditioned fear. SP, however, never demonstrated a SCR to the blue square, even though her SCR to the mild shock was normal. When shown her data indicating a lack of conditioned fear, SP commented:

> I knew that there was an anticipation that the blue square, at some particular point in time, would bring in one of the volt shocks. But even though I knew that, and I knew that from the very beginning, except for the very first one when I was surprised. That was my reaction-I knew it was going to happen. So I learned from the very beginning that is was going to happen: blue-shock. And it happened. I turned out to be right! (Phelps, 2002, p. 559).

Thus, in spite of the lack of a CR as assessed with a physiological measure, SP had explicit awareness and understanding of the events of fear conditioning. Patients with damage to the hippocampus, whose amygdalas are intact, show the opposite pattern (Bechara et al., 1995). That is, they have a normal CR as measured implicitly through physiological arousal, but they are unable to report explicitly the events of fear conditioning. This dissociation between an explicit understanding of the emotional properties of the CS and an implicit assessment of the emotional significance of the CS suggests that the mechanisms of explicit and implicit emotional evaluation rely on distinct neural circuits. This discovery has obvious implications for dissociations that

are observed in the behavioral responses of humans to complex situations in which some behaviors, such as thoughts and feelings, can be disjointed from action.

From Animal Models of Fear Conditioning to the Neural Systems of Race Bias

In particular, animal data on fear conditioning indicating a dissociation between implicit and explicit responses are reminiscent of findings from humans that illustrate several dissociations. Group data show small to no preferences on explicit measures of attitudes, whereas implicit measures of attitudes reveal large preferences. Or, explicit attitudes demonstrate large preferences that are not mimicked on implicit measures of the same. For example, white Americans express small to no preference (depending on the type of sample) for their own group compared to African-Americans. In the implicit association test (IAT), subjects see items that belong to one of four categories: photos of Black and White individuals and words for good and bad concepts (joy, love, friend versus devil, vomit, agony). For half the trials, the task involves classifying White + good using one computer keyboard key while also classifying Black + bad using another key. For the remainder of the trials, the opposite pairing is requested, with classification on White + bad using a single key and Black + bad using another. The difference in response latencies in these two conditions constitutes the IAT effect and is regarded to be a measure of evaluative strength of association regarding these race groups. Responses on the IAT (Greenwald & Banaji, 1995) using large numbers of White American subjects showed stronger association of White + good-Black + bad than to Black + good-White + bad.

On the other hand, African-Americans verbally report strong preference for their own group compared to white Americans, but their IAT data show a much weaker preference for their own group compared to that of white Americans. About half of African-Americans, as opposed to 72 percent of whites, favor their own group.

Animal data on conditioning had led to hypotheses concerning humans with damage to either the hippocampus or amygdala. Similarly, these models offered a starting point to infer the mechanisms that underlie the dissociations in attitudes observed in normal humans. Moreover, human data on implicit-explicit dissociations in race and other group attitudes had raised doubts about the validity of the measure. What is the meaning of the associative strength observed in response latencies that reveal faster or slower responses to some pairings over others? Many objections were raised (the measure reflects an effect of familiarity, not attitude; the measure taps something that has no predictive validity), but the one for which animal models proved particularly useful concerned a specific objection regarding the construct validity of the IAT.

The claim was that the IAT response was not a measure of evaluation or affect at all; in other words, it was not a measure of a preference or attitude. It was, instead, a measure of colder associations that are not typically viewed as revealing an attitude-a warm if not hot construct that was devised to capture the evaluative or affective portion of one's responses to any object.

The animal data and its convergence with patient and fMRI data provided the setting to conduct the first study of this sort to investigate the construct validity of the IAT while also shedding light on the dissociation itself. To test this hypothesis, we investigated whether the amygdala, which mediates implicit physiological indications of conditioned fear, might also mediate implicit expressions of race bias as measured by both the IAT and a physiological assessment (Phelps et al., 2000). To the extent that this is observed, behavioral data resulting from the IAT can be assumed to be tapping into an affective process.

Phelps et al. (2000) used fMRl to examine amygdala activation in White Americans while they observed pictures of Black and White unfamiliar male faces and correlated observed brain activity with two behavioral measures, the IAT and startle eyeblink. During imaging, subjects were asked to indicate by button press whether a face was repeated. Afterward, they were given the two implicit measures of race bias; the IAT, a reaction time measure of cognitive conflict, and startle eyeblink, a physiological assessment of evaluation, as well as an explicit measure of race bias, the Modern Racism Scale (MRS; Mcconahay, 1986). The IAT effect was computed as a difference score between the speed to respond to White + good-Black + bad pairs and the opposite, White + bad-Black + good pairs. Social groups were represented using faces that could be clearly classified as belonging to one or the other group. Good and bad stimuli were words that again, could be clearly classified as good or bad (love, joy, friend versus agony, vomit, devil).

The physiological assessment was a measure of the startle reflex, a natural response to a startling stimulus, such as a loud noise. One of the first components of this reflex response is an eyeblink, the strength of which can be assessed in the laboratory by measuring the reaction of muscles around the eyes. Startle responses are enhanced or potentiated in the presence of negative stimuli relative to neutral or positive stimuli (Lang, Bradley, & Cuthbert, 1990). In our assessment of implicit race bias, subjects were startled with a loud white noise while viewing pictures of unfamiliar Black and White male faces. The difference in strength of the startle eyeblink response while viewing Black versus White faces was the measure of race bias.

Variability among the White American subjects was seen in the amygdala response to Black versus White faces. Although most subjects showed greater amygdala activa-

tion while viewing black compared to white faces, the overall effect was not significant. We had expected overall greater amygdala activation to Black than to White faces, but the more important idea was the relationship between the magnitude of amygdala activation and the IAT, startle reflex, and MRS. When amygdala activation to Black versus White faces was correlated with the three measures of race bias, only the implicit measures predicted amygdala activation. On region of interest analysis, amygdala activation correlated significantly with implicit race bias as measured by the IAT ($r = 0.576$) and startle eyeblink ($r = 0.556$), with more negative assessments of Black relative to White on the measure of behavior predicting greater amygdala activation. No correlation was seen between amygdala activation and the MRS. An additional exploratory analysis generated correlation maps indicating brain regions where behavioral measures of race bias were significantly correlated with the strength of activation to black versus white faces. These correlation maps indicated that a region of the right amygdala correlated with both the IAT and startle eyeblink measures of bias (figure 12.1, plate 5).

This study revealed a variety of results that can be the basis of future research. First, learning what a social group means, along the evaluative dimension, involves mechanisms that are similar across species. For example, in Black Americans, greater amygdala activation was evident to White versus Black unfamiliar male faces (Hart et al.,

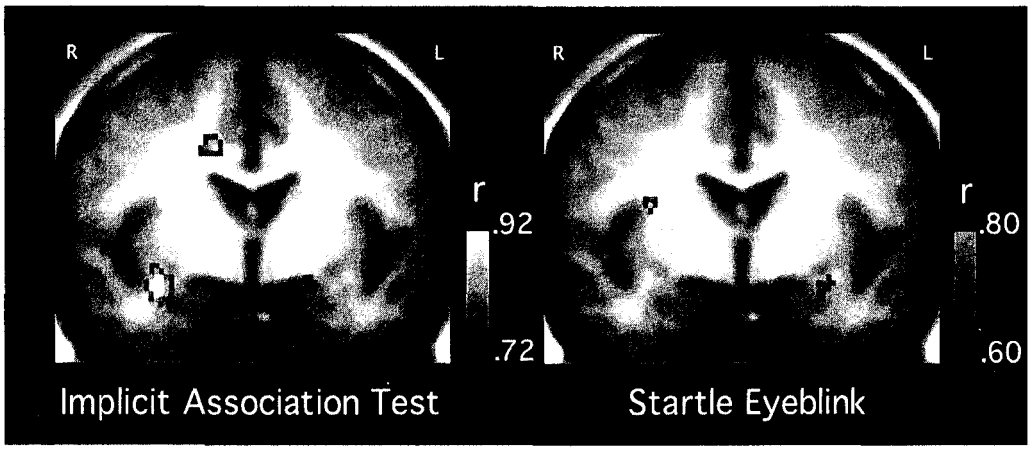

Figure **12.1**
Correlations maps indicating regions where the differential BOLD response to Black versus White was significantly correlated with the strength of implicit race bias as assessed by the IAT (left) and startle eyeblink (right). (Reprinted from Phelps et al., 2000.) See plate 5 for color version.

2000), suggesting that this effect extends to outgroup faces in general. Another study observed amygdala activation in White Americans to Black versus White faces that are presented subliminally, so quickly that subjects are unaware of their presentation and report seeing only a flicker (Cunningham et al., 2004).

The amygdala is implicated in learning the emotional significance of an event, such as fear. It also is implicated in fear extinction, in which a CS is no longer paired with a UCS and, eventually, the CS fails to elicit a CR. During initial extinction learning, the amygdala appears to play an active role in learning that the CS no longer predicts an aversive event that it previously did (Falls, Miserendino, & Davis, 1992; Phelps et al., 2004). That is, after a period of time of no pairing, the amygdala response to an extinguished CR is diminished, and other brain regions, such as the prefrontal cortex (PFC), appear to play a role in the retention of extinction learning and inhibiting an amygdala response (Morgan, Romanski, & LeDoux, 1993; Phelps et al., 2004; Quirk et al., 2000).

Jn terms of race bias, we might expect two types of extinction learning. The first is extinction to specific individual members of a race group who are familiar. A further question of interest concerns the extent to which such effects with specific individuals generalize beyond those instances to include new members of Black or White race categories who were not previously involved in the "learning" of extinction.

To assess the effect of familiarity and exposure on the neural systems of race bias, we repeated the fMRI study described above, with one exception. The Black and White male faces were not unfamiliar, but rather were well known figures from the popular media with relatively positive public images (Martin Luther King, John F. Kennedy, Denzel Washington, Harrison Ford). With these stimuli, we failed to find consistent amygdala activation. Much like the lack of amygdala activation to CSs that had been previously extinguished (Phelps et al., 2004), the amygdala response in White subjects did not differentiate same and other race group faces. There was also no accompanying correlation between amygdala activation and implicit or explicit measures of race bias (Phelps et al., 2000), suggesting that individuals who are well known may be subtyped so as not to evoke favorable-unfavorable feelings themselves, whereas the social category as whole continues to elicit them.

In the Phelps et al. (2000) study, we did not explore responses to the entire brain, so we do not know if the PFC was involved in retention of this extinction learning, or inhibition of the amygdala response. However, Cunningham et al. (2004) found that the magnitude of activation in regions of the PFC is accompanied by lower activity in the amygdala in the Black versus White comparison, similar to extinction learning (Phelps et al., 2004). However, this was observed only when subjects were aware

of the presentation of Black and White faces, suggesting a consciously mediated mechanism of cognitive control that may be specific to humans. In addition, Richeson et al. (2003) reported that activation of the PFC may be related to the inhibition of race bias. The precise regions of the PFC observed in these studies differ from those observed in extinction learning (Phelps et al., 2004), but they are consistent with studies of emotion regulation (Ochsner et al., 2002), suggesting that cognitive control of emotion, like extinction learning, results in the interaction of the PFC and amygdala. Although not able to speak directly to the issue of communication between subcortical and cortical regions, these studies suggests the potential that conscious values and beliefs may have an impact on the subcortical mechanisms of race bias. They speak to the question of the role of individual differences in conscious attitudes and, in this case, the ability to modulate the negative response to Black versus White race groups through the work of mechanisms that are more directly under conscious control.

Although we have not explored how the amygdala response to all members of another race group (familiar and unfamiliar) might change with extensive positive exposure to a number of individuals from that group, it was suggested that intergroup contact is a factor that predicts positive outgroup attitudes and interactions (Pettigrew & Tropp, 2000). These results are consistent with the hypothesis that exposure, without aversive consequences, may help extinguish culturally acquired race bias.

From Animal Models of Fear Conditioning to Social Learning

Thus far, we have argued that the neural systems of classical fear conditioning might be similar to the neural systems of culturally acquired race bias. However, a fundamental difference between conditioned fear and race bias is the means of acquisition. In fear conditioning, the subject learns the emotional significance of a CS through direct, personal aversive experience (UCS) paired with the CS. Culturally dependent race bias is learned through social means, that is, interaction and communication between individuals, however subtle the exchanges are. Although on rare occasions race bias may be acquired through aversive experience, in most cases, it is learned without personal, aversive consequences. Given this, it is important to show that social means of fear learning rely on the same neural mechanisms as learning through direct aversive experience.

A primary means of social learning in humans is language-a symbolic form of communication between individuals. Studies on the acquisition of phobias, suggest that linguistic communication aboul Lhe potential aversive consequences of stimuli can

generate potent fears (King, Eleonora, & Ollendick, 1998). For example, individuals suffering from a phobia of germs have extreme reactions to stimuli they have been told about-microscopic organisms they have neither seen nor will see. Some individuals go to great lengths to avoid these situations. An experimental paradigm to investigate the verbal communication of fears is instructed fear. This is similar to fear conditioning, except that subjects are simply told that they might receive a shock paired with a stimulus (the Instructed CS). No shocks are ever actually delivered. Instructed fear results in similar physiological expressions of fear as actual fear conditioning involving an aversive stimulus (Hugdahl & Ohman, 1977).

There is little reason to expect that the amygdala would play a significant role in the acquisition of verbally acquired fears. Simply being told that a neutral stimulus may, in the future, predict a possible aversive consequence is unlikely to elicit a strong emotional response at the time this information is learned. However, it is possible that the expression of these verbally acquired fears may rely on similar mechanisms as fears acquired with direct aversive experience with fear conditioning. In an effort to determine if the amygdala is involved in verbally acquired fears, we investigated the neural systems of instructed fear using fMRI (Phelps et al., 2001). Subjects were informed that they would see several presentations of a blue square, and that paired with some of these presentations they would also receive a mild shock to the wrist. None of the subjects actually received a shock at any time. Afterward, all subjects indicated awareness that a shock would be presented with a blue square. Consistent with this, the SCR response to the blue square was increased, a physiological indication of fear. Similar to fear conditioning, there was significant activation of the left amygdala when subjects believed they might receive a shock (figure 12.2, plate 6). As with the earlier study with fear conditioning and implicit measures of race bias, the strength of amygdala activation was correlated with the strength of the SCR response to the blue square.

These results suggest that the amygdala responds to verbally instructed fears, but brain imaging data cannot indicate if the amygdala plays a critical role in the expression of these fears. This was shown in a second study in patients with amygdala damage (Funayama et al., 2001). Using a similar paradigm, there patients failed to show a physiological indication of instructed fear, although they were able to report verbally that the blue square indicated the possibility of shock. Although studies examining the neural systems of instructed fear differ from fear conditioning in that only the left amygdala is involved, perhaps because of the verbal nature of the learning and representation, in both paradigms the amygdala was critical for the

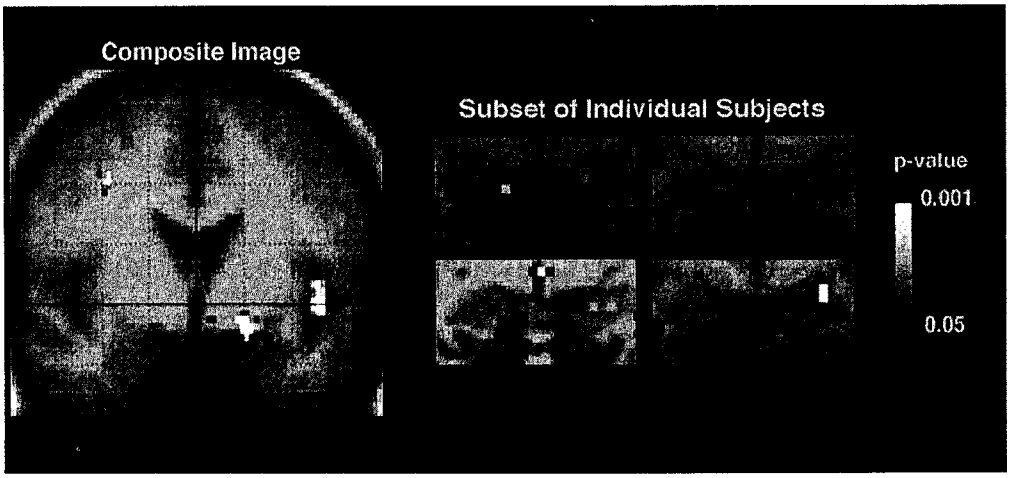

Figure 12.2
Activation to instructed fear. A group composite map (left) and selected individual suhjects (right) indicate regions with a significantly greater BOLD response to a "threat" stimulus (paired with possibility of shock) versus a "safe" stimulus (no possibility of shock). (Reprinted from Phelps et al., 2001.) See plate 6 for color version.

physiological expression of fear learning. These studies indicate that fears acquired through social symbolic means, that are imagined and anticipated but never actually experienced, depend on similar mechanisms as fears acquired through direct, aversive experience.

Another powerful means of social lea rning is observation (Bandura, Ross, & Ross, 1961; Hygge & Ohman, 1978). By watching other individuals, we can lea rn about situations that should be approached or avoided ourselves. Emotional learning through observation has been demonstrated in infants viewing the emotional reactions of their mothers (Gerull & Rapee, 2002), as well as in monkeys that observed reactions of other monkeys to specific stimuli (Mineka & Cook, l993). A recent study explored the physiological expression of fear acquired through social (instruction and observation) and nonsocial (classical conditioning) means (Olsson & Phelps, 2004). In the observation condition, subjects watched a video of another individual responding to a stimulus (observa tional CS) that was paired with a mild shock to the wrist (observational UCS). They were told they would experience a similar procedure immediately after viewing the video. The subjects were then presented the observational CS, but never actually received a shock. All three methods of fear learning resulted in similar levels of

physiological fear expression when the stimulus (conditioned, instructed, or observational CS) that was paired with the aversive event was presented supraliminally, so that subjects were aware of its presentation (Olsson & Phelps, 2004). Of interest, both fear conditioning and observational fear resulted in the physiological expression of fear when the CS was presented subliminally-so quickly subjects were unaware it was presented. When an instructed CS was presented subliminally, there was no indication of a fear response (Olsson & Phelps, 2004). Thus, observing another individual's response to a CS paired with an aversive UCS results in a representation that is similar to direct experience with the aversive event.

The similarity in expression of fear to a subliminal CS in fear conditioning and observational fear suggests overlapping neural circuits. Although the acquisition of verbally instructed fears is unlikely to involve the amygdala, it is possible that the vicarious experience of observing another individual's aversive reaction to a UCS paired with a CS evokes an amygdala response. To investigate this, an fMRI study (Olsson et al., 2004) examined amygdala activation when observing a confederate receiving a mild shock paired with a CS (acquisition stage) and then participating in the same study (test stage), although shocks were never actually delivered. Significant amygdala activation occurred to both observing a confederate receive a shock and expecting to receive shock oneself. Thus the amygdala may be involved both in the acquisition and expression of observational fear learning, similar to fear conditioning.

Studies on the neural systems of instructed and observational fear learning indicate that socially acquired negative evaluations may not be fundamentally different than those acquired through classical conditioning. The amygdala plays a role in the acquisition (observation) and expression (observation and verbal instruction) of fears learned through social means. Given this, one might expect culturally acquired preferences, which are learned through social communication and observation, to rely on similar neural mechanisms as classical conditioning, as our previous studies suggested.

From Animal Models of Fear Conditioning to the Dynamics of Race Bias and Learning

If classical fear conditioning is an appropriate model for socially acquired race bias, we should be able to use this model to generate novel predictions concerning responses to race group stimuli. Previous studies of classical conditioning indicated that not all CSs are created equal. Some stimuli, when paired when an aversive event, will result in stronger expression and slower extinction of a CR. These stimuli have been called "prepared," to capture the idea that they are evolutionarily prepared to

support fear learning. For instance, in both humans and monkeys, when a stimulus representing a snake (such as a picture or a toy snake) is paired with a UCS it will quickly acquire a CR that is stronger and less prone to extinction than a CS representing something more neutral, such as a flower or mushroom (see Ohman & Mineka, 2001, for a review). Although it may seem that more potent CRs to stimuli representing snakes, which already have a negative connotation for most people, may simply reflect a previously acquired negative evaluation, two lines of evidence suggest that this b not the case. Monkeys raised in the laboratory that never had experience with snakes will show a prepared CR to a toy snake through fear conditioning or observational fear learning (Mineka et al., 1984).

In humans, discrimination fear conditioning paradigms, in which one stimulus (CS+) is paired with the UCS and another stimulus (CS-) is not and serves as a baseline, suggest that instances of some categories can serve as prepared stimuli. A specific representation of a snake when used as a CS+ will show a prepared CR, even when the baseline is responses Lo another representation of a sqake that is used as a CS-; that is, one that was not paired with an additional aversive stimulus. Even if a class of stimuli has an overall negative evaluation, a specific instance of that class, when used as a CS and paired with aversive consequences, will acquire a CR that is stronger and more resistant to extinction than a specific CS from a class of stimuli that is not prepared (Ebert et al., 2005). We asked to what extent can social groups that vary in race or ethnicity be such prepared stimuli?

Jn an effort to examine if the race of a face (as a CS) will alter the characteristics of a CR, we examined fear conditioning to pictures of White and Black unfamiliar males faces in White and Black American subjects. We hypothesized that an outgroup face, when used as a CS, would result in a CR that was more resistant to extinction than an ingroup face CS. A discrimination fear conditioning paradigm was used in which there were four CSs, two pictures of Black faces and two of White faces. One picture from each race group served as a CS+ and was paired with a mild shock to the wrist (UCS) during acquisition, while the other served as a CS- and was never paired with shock. For each race group, we compared the SCR to the CS+ to the same race CS- as an indication of fear conditioning. Regardless of subject's race, both Black and White face CS+ stimuli showed evidence of acquisition of a CR. However, the CR to an outgroup did not show the same rate of extinction as the CR to an ingroup face. In other words, an outgroup race face when paired with an aversive event is resistant to extinction learning, indicating that fear learning to a specific member of an outgroup race will be less likely to diminish than fear learning to a specific member of an ingroup race (Ebert et al., 2005). These resulls parallel those obtained when comparing

extinction to CRs to specific instances of snakes versus specific instances of flowers. In fear conditioning, outgroup race faces look like prepared stimuli.

The difference in rate of extinction to another race face CS suggests that a learned negative evaluation to a specific member of an outgroup race may be especially difficult to change, even with additional exposure and familiarity with no aversive consequences. Although exlinction learning may help diminish negative bias to outgroup race members that have not been specifically linked to aversive consequences, if an outgroup race member is linked to a negative event, this learned response appears to be more durable than an acquired negative response to an ingroup race member. Using classical fear conditioning as a model, we have been able to generate novel predictions about the impact of race group membership on differential levels of emotional learning.

Conclusion

As the research outlined in this chapter indicates, animal models can be quite useful in understanding a wide range of human behaviors, including complex social interactions. In our studies we relied on models of conditioned fear as a starting point in our investigations of the neural systems of race bias and even used these models to derive direct hypotheses of social learning and race bias. In this sense, animal models are useful in that they provide a way to detect those aspects of social learning and memory that may be extensions of the processes that are common to all animals. On the other hand, what makes humans *human* is the ability for conscious awareness, intention, and control to shape behavior. A human being, unlike other animals, can regulate behavior from deciding to delay eating gratification, even when hungry, to dine with a friend, to deciding to recode stimuli that produce negativity into those that are neutral or positive. As studies on the conscious control of race bias (Cunningham et al., 2004; Richesonn et al., 2003) and the cognitive control of emotion (Oschner et al., 2002) indicate, different neural mechanisms within the PFC may be specifically linked to these human characteristics. How these human mechanisms interact with more general neural systems of emotion inhibition that are common across species (Phelps et al., 2004) is unclear. It is clear, however, that animal models can be useful tools in our investigations of the neural systems underlying human social behavior. They serve as the starting poi nt from where we can learn about what links us to other animals and distinguishes us from them.

Acknowledgments

Supported by grants from the NIMH and the McDonnell Foundation to E. A. Phelps, and NIMH and Rockefeller Foundation's Bellagio Study Center to M. R. Banaji. We thank Molain Saintilus for help preparing the manuscript.

References

Band ura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression th rough imitation of aggressive models. *journal of Abnormal Social Psychology,* 63, 575-582.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Oamasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in human. *Science,* 269(5227), 1115-1118.

Buchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron,* 20(5), 947-957.

Cahill, L., Wein berger, N. M., Roozendaal, B., & McGa ugh, J. L. (1999). Is the amygdala a locus of "conditioned fear"? Some 4uestiuns and caveats. *Neuron,* 23(2), 227-228.

Cohen, J. D., Noll, D. C., & Schneider, W. (1993). Functional magnetic resonance imaging: Overview and methods for psychological research. *Behavioral Research Methods, Instruments and Computers,* 25(2), 101-113.

Cu nningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., *Gore, J.* C., & Bana ji, M. R. (2004). Separable neural components i n the processing of black and white faces. *Psychological Science, 15,* 806-813.

Davis, M. (1997). Neurobiology of fear responses: The role of the amygdala. *Journal of Neuropsychiatry and Clinical Neuroscience,* 9(3), 382-402.

Ebert,]., Olsson, A., Phelps, E. A., & Ba na ji, M. R. (200.). Classical conditioning effects during extinction as a measure of race bias. Poster presented at the meeting of Society for Personality and Social Psychology.

Palls, W. A., Miserendino, M. J., & Davis, M. (1992). Extinction of fear-potentiated startle: Blockade by infusion of an NMDA antagonist into the amygdala. *Journal of Ne111mcience, 12(3),* 854-863.

Funayama, E. S., Grillon, C., Davis, M., & Phelps, E. A. (2001). A double dissociation in the affective modulation of startle i n humans: Effects of u nilateral temporal lobectomy. *journal of Cognition Neuroscience, 13(6),* 721-729.

Gerull, F. C. & Rapee, R. M. (2002). Mother knows best: Effects of maternal modelling on the acquisilion of fear and avoidance behaviour in toddlers. *Behaviour Rcscard1 and Therapy,* 40(3), 279-287.

Greenwald, A. G. & Bana ji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review,* 102(1), 4-27.

Hart, A. *J.,* Whalen, P.]., Shin, L. M., Mcinerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *NeuroH'port,* 11(11), 2351-2355.

Hugdahl, K. & Ohman, A. (1977). Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *Journal of Experimental Psychology,* 3(5), 608-618.

Hygge, S. & Ohman, A. (1978). Modeling processe in the acquisition of fears: Vica rious electrodermal conditioning to fear-relevant tim uli. *fo111nal Personal Social Psychology,* 36(3), 271-279.

Kapp, B. S., Frysinger, R. C., Gallagher, M., & Haselton,]. R. (1979). Amygdala central nucleus lesions: Effect on heart rate conditioning in the rabbit. *Physiology and Behavior,* 23(6), 1109-1117.

King, N.]., Eleonora, G., & Ollendick, T. H. (1998). Etiology of childhood phobias: Current status of Rachman's three pathways theory. *Behavior Research and TherafJ)',* 36(3), 297-309.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, j. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron,* 20(5), 937-945.

LaBar, K. S., LeDoux, **J.** E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fea r conditioning following u nilateral temporal lobectomy in huma ns. *Journal of Neuroscience,* 15(10), 6846-6855.

Lang, P.]., Bradley, M. M., & Cuth bert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review,* 97(3), 377-395.

LeDoux, J. E. (1996). *The Emotional Brain.* New York: Simon & Schuster.

LeDoux, J. E. (2002). *Synaptic Se/(" How Our Brains Become Whu We Are.* New York: Viking.

Mcconahay, J. P. (1986). Modern racism, ambivalence, and the modern racism scale. Jn J. F. Dovidio & S. L. Gaertner (Eds.), *Pre;udice, Discrimination and Racism* (pp. 91-125). Orlando: Academic Press.

Mineka, S. & Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimenlul Psyclrology,* J22(1), 23-38.

Mineka, S., Davidson, M., Cook, M., & Keir, R. (1984). Observational conditioning of sna ke fear in rhesus monkeys. *Journal of Abnormal Psychology,* 93(4), 355-372.

Morgan, M. A., Romanski, L. M., & LeDoux, **J.** E. (1993). Extinction of emotional learning: Contribution of medial prefrontal cortex. *Neuroscience Letter, 163,* 109-1 B.

Ochsner, K. N., Bunge, S. A., Gross,].]., & Gabrieli, J. D. (2002). Rethinking feelings: An tMI{] study of the cognitive regulation of emotion. *Journal of Cognitive Ne11roscience, 14(8),* 1215-1229.

Ohman, A. & Mineka, S. (2001). Fea rs, phobias, and preparedness: Toward a n evolved module of fear and fear learning. *Pycl10logical Review,* 108(3), 483-522.

Olsson, A., Nearing, K., Zheng, J., & Phelps, E. A. (2004). Learning by observing: Neural correlates of fear learning through social observation. Presented at the 34<h annual meeting uf the Society for Neuroscience, San Diego.

Olsson, A. & Phelps, E. A. (2004). Learned fear of "unseen" faces. *Psychological Science, 15,* 822-828.

Pettigrew, T. F. & Tropp, L. R. (2000). Does intergroup contact reduce prejudice: Recent meta-analytic findings. In S. Oskamp (Ed.), *Reducing Prejudice and Discrimination* (pp. 93-114). Mahwah, NJ: Erlbaum.

Phelps, E. A. (2002). The cognitive neuroscience of emotion. In M. S. Gazzaniga, R. B. Ivry, & G. R. Magnum (Eds.), *Cognitive Neuroscience: The Biology of M ind,* 2nd ed. (pp. 537-576). New York: Norton.

Phelps, E. A., Delgado, M. R., Neari ng, K. l., & LeDoux, J. E. (2004). Extinction learning in h umans: Role of the amygdala and vmPFC. *Neuron, 43,* 897-905.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., l:'unayama, E. S., Gatenby, J . C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *journal of Cognitive Neuroscience,* 12(5), 729-738.

Phelps, E. A., O'Connor, K. J., Gatenby, *J.* C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nature Neuroscience,* 4(4), 437-441.

Quirk, G. J., Russo, G. K., Ba rron, J. L., & Lebron, K. (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *Journal of Neuroscience,* 20(16), 6225-6231.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & Shelton, J. N. (2003). An fMRl investigation of the impact of interracial contact on executive function. *Journal of Ne11roscie11ce,* 6, 1323-1328.