Behavioral/Cognitive

Diagnostic Value Underlies Asymmetric Updating of Impressions in the Morality and Ability Domains

Peter Mende-Siedlecki,[1] Sean G. Baron,[1] and Alexander Todorov[1,2]

[1]Department of Psychology, Princeton University, Princeton, New Jersey 08542 and [2]Behavioural Science Institute, Radboud University, 6500 HE, Nijmegen, The Netherlands

While positive behavioral information is diagnostic when evaluating a person's abilities, negative information is diagnostic when evaluating morality. Although social psychology has considered these two domains as orthogonal and distinct from one another, we demonstrate that this asymmetry in diagnosticity can be explained by a single parsimonious principle—the perceived frequency of behaviors in these domains. Less frequent behaviors (e.g., high ability and low morality) are weighed more heavily in evaluations. We show that this statistical principle of frequency-derived diagnosticity is evident in human participants at both behavioral and neural levels of analysis. Specifically, activity in right ventrolateral prefrontal cortex increased preferentially when participants updated impressions based on diagnostic behaviors, and further, activity in this region covaried parametrically with the perceived frequency of behaviors. Activity in left ventrolateral PFC, left inferior frontal gyrus, and left superior temporal sulcus showed similar patterns of diagnosticity and sensitivity, though additional analyses confirmed that these regions responded primarily to updates based on immoral behaviors.

Introduction

When evaluating other people, we are faced with the task of accurately assessing two key components of character—are this person's intentions bad or good, and can they follow through on those intentions? These evaluations map onto the dimensions of morality (e.g., incorporating "warmth," "trustworthiness," and "kindness") and ability (e.g., incorporating "competence," "intelligence," and "efficacy"), and have been extensively studied as universal, orthogonal dimensions of person perception (Fiske et al., 2007).

Behavioral research suggests that learning about other people in the domains of morality and ability is characterized by a asymmetry in diagnosticity (Skowronski and Carlston, 1987; Wojciszke, 2005). Specifically, immoral actions receive more weight than moral actions (Reeder and Spores, 1983; Crocker et al., 1984; Reeder and Coovert, 1986), while competent actions receive more weight than incompetent actions (Reeder et al., 1977; Reeder and Fulks, 1980, Skowronski and Carlston, 1987; Kubicka-Daab, 1989; Brycz and Wojciszke, 1992; Lewicka et al., 1992; Wojciszke et al., 1993). Various frameworks have attempted to explain these asymmetries, (Reeder and Brewer, 1979; Skowronski and Carlston, 1989); however, one parsimonious statistical principle can potentially bind these accounts together. Many of these lines of research invoke the statistical frequency of behavior, stating, to some extent, that behaviors perceived as being rare should have more influence on evaluations (Fiske, 1980; Reeder, 1993; Wojciszke et al., 1993). Potentially, high-ability and low-morality behaviors are less common in the environment than their cross-valence counterparts, or at the very least, people's perceptions of behavioral base rates reflect this pattern. While this has yet to be empirically tested, a connectionist model of impression formation built upon this assumption robustly replicates both the negativity bias in the morality domain and the positivity bias in the ability domain (Van Overwalle and Labiouse, 2004), suggesting that the statistical frequency (either experienced or perceived) of a given behavior might weight the updating of another person's character disposition on the basis of that behavior.

Despite extensive behavioral research, social neuroscience has yet to address distinctions between the ability and morality domains. The past decade has yielded many neuroimaging studies on behavior-based impression formation (Mitchell et al., 2004, 2005, 2006; Schiller et al., 2009; Freeman et al., 2010; Cloutier et al., 2011a), which either focused exclusively on the morality domain or collapsed across multiple domains. Ensuing research has explored the neural dynamics supporting impression updating in light of inconsistent or expectancy-violating behavioral information (Baron et al., 2011; Cloutier et al., 2011b, Ma et al., 2012; Ames and Fiske et al., 2013; Mende-Siedlecki et al., 2013), but despite strong convergence, the principles underlying the resolution of these inconsistencies are unclear.

Across two studies, we examined the roots and impacts of behavioral diagnosticity in the context of impression updating. In Study 1, we sought behavioral evidence that the diagnostic asymmetry between the ability and morality domains is linked to differing perceptions of the frequency of behaviors. In Study 2, we examined the neural dynamics underlying updating based on

Received May 31, 2013; revised Sept. 21, 2013; accepted Oct. 25, 2013.

Author contributions: P.M.-S., S.G.B., and A.T. designed research; P.M.-S. and S.G.B. performed research; P.M.-S. analyzed data; P.E.M.-S. and A.T. wrote the paper.

This research was supported by National Science Foundation (NSF) Grant BCS-0823749 to A.T. and NSF Grant DGE 1148900 to P.M.-S. We thank Rui Costa, Elizabeth Carnahan, and Ashley Halkett for their assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Peter Mende-Siedlecki, Psychology Department, Princeton University, Princeton, NJ 08540. E-mail: pmende@princeton.edu.

DOI:10.1523/JNEUROSCI.2334-13.2013

0270-6474/13/3319406-10$15.00/0

ability versus morality information. We predicted that behavioral diagnosticity would guide updating-related responses on behavioral and neural levels.

Materials and Methods

Study 1

First, we directly tested whether people have different expectations about the base rates of behaviors, as a function of their valence and domain. Specifically, participants were asked to estimate the frequency of individual behaviors varying on ability and morality.

Participants. Eighty participants (47 female, 33 male; mean age, 33.72 years; age range, 18 –75 years) were recruited from the Amazon Mechanical Turk site and were paid $0.50 for their participation. We acquired informed consent for participation approved by the Institutional Review Board for Human Subjects at Princeton University and debriefed participants at the completion of the experiment.

Stimuli and procedure. Participants were asked to consider a series of behaviors taken from a compendium of 400 behaviors previously rated on kindness and intelligence (Fuhrman et al., 1989). Based upon these ratings, we selected 200 behaviors that specifically implied moral character (either highly moral or highly immoral behaviors) or ability (either highly competent or highly incompetent behaviors), based on ratings of kindness and intelligence, respectively. Previous research in this area has typically attempted to balance out positive and negative behaviors on relevant dimensions (e.g., competence, honesty) with respect to their distance from the midpoint of the dimension (Skowronski and Carlston, 1987; Wojciszke et al., 1993).

For the stimulus set at hand, we also needed to be sure that our selected 200 behaviors were not biased to yield an asymmetric subset of behaviors from the overall sample. Specifically, we assessed whether the selected behaviors differed in terms of extremity, comparing between moral and immoral, and competent and incompetent behaviors, separately. To do so, we normalized the ratings of the selected behaviors based on the original 400-behavior sample, and tested for differences between positive and negative behaviors within each domain.

We capitalized on a cluster analysis performed by Fuhrman et al., 1989. To determine the overall mean for morality behaviors, we averaged across kindness ratings of all behaviors from clusters relevant to the morality domain, while for ability behaviors we averaged across intelligence ratings of all behaviors from clusters relevant to the ability domain. (We chose not to compute one grand mean across the 400-behavior set to avoid spillover effects for a number of behaviors, which, though classified by the cluster analysis primarily indicating high morality or low morality, were also rated very extremely on ability. Moreover, since we did not have concrete predictions regarding the sixth cluster—essentially neutral behaviors—we excluded it from both averages.) We used these two means to normalize the selected behaviors from the ability and morality domains, to compare the extremity of positively and negatively valenced behaviors.

In the morality domain, moral behaviors were significantly more extreme than immoral behaviors ($p < 0.001$; immoral behaviors = 3.165 vs moral behaviors = 3.506). In the ability domain, competent and incompetent behaviors did not differ significantly in terms of extremity ($p = 0.218$; incompetent behaviors = 2.413 vs competent behaviors = 2.600). These data suggest that the behaviors we selected were not intrinsically biased in the direction of the predicted asymmetries in the ability and morality domains, at least in the context of the larger set of behaviors from Fuhrman et al., 1989. (While we did observe a significant difference in the extremity of the moral and immoral behaviors we selected, this difference was in the opposite direction of the expected negativity bias in the morality domain and, as such, would have worked against the predicted effects.)

Due to time constraints, each participant saw a random subset of 100 behaviors taken from the larger set of 200. For each behavior, participants were asked how many people of a random sample of 100 would have performed that behavior at some point in time.

Study 2

Next, we devised a paradigm based on our previous research (Mende-Siedlecki et al., 2013), in which participants learned about a series of individuals and were occasionally presented with information that may cause them to update their impressions of these individuals. In terms of the behavioral data, we tested (1) whether behavioral ratings of trustworthiness and competence show evidence of diagnostic asymmetries, and (2) whether perceived frequency (as collected in Study 1) predicts changes in behavioral ratings. Furthermore, we tested whether this principle of frequency-derived diagnosticity guides impression updating on a neural level as well. If so, the diagnosticity of behavioral information, rather than its content (morality vs ability) may drive neural responses associated with impression updating. More importantly, we should observe a relationship between the neural correlates of diagnosticity-specific updating and regions whose activity tracks the relative frequency of behaviors. We also tested for valence effects, in an attempt to isolate regions showing preferential activity when updating is based on negative behaviors as opposed to positive behaviors.

The diagnosticity account is not without alternative possibilities, of course. On one hand, if learning in the ability and morality domains is truly distinct, and dependent on separate inferential processes (Reeder, 1993, 2006), one might expect updating based on ability information to bear a different neural signature from updating based on morality information. On the other hand, given the preponderance of work suggesting that affectively negative stimuli outweigh their positive counterparts both in general (Baumeister et al., 2001) and in the context of impression formation (Wojciszke et al., 1998; De Bruin and van Lange, 2000), one might expect updating based on negative behavioral information (i.e., highly immoral or highly incompetent behavior) to have the strongest impact on activity in regions involved in updating impressions. In Study 2, we test both of these alternative hypotheses.

Participants. Twenty-three participants (13 female), ages 18 –31 years (mean age, 22.5 years) volunteered for the fMRI study and were paid $30 for their participation. All participants were right handed, had normal or corrected-to-normal vision, and reported no history of neurological illnesses or abnormalities. We acquired informed consent for participation approved by the Institutional Review Board for Human Subjects at Princeton University and debriefed participants at the completion of the experiment.

Stimuli. Each participant saw a series of 50 male and female faces taken from the Karolinska Directed Emotional Faces set (Lundqvist et al., 1998) paired with sets of behaviors constructed from the stimuli tested in Study 1. Each individual face was presented in five consecutive trials, each time with a different sentence describing a behavior that the individual had performed. Each of the five consecutively viewed behaviors varied within either the ability or the morality domain. Critically, for each individual, the valence of behavior switched on the fourth behavior. For instance, a person who was previously presented as very competent and capable might suddenly be shown in a more incompetent light (Fig. 1). Our primary rationale behind this design was to instantiate a strong expectation of how a given individual would behave over time, rather than presenting simple trial-to-trial inconsistencies. This design was developed and validated over several iterations of behavioral piloting and was used in a previous neuroimaging investigation of impression updating (Mende-Siedlecki et al., 2013).

The same 200 behaviors used in Study 1 were selected for use in Study 2 and divided into five-behavior groups. We reiterate that these behaviors were initially selected based upon kindness and intelligence ratings (Fuhrman et al., 1989), and that our selections were not intrinsically asymmetric, compared with the original 400-behavior sample.

Ability individuals consisted of faces paired with either three consecutive competent behaviors, followed immediately by two incompetent behaviors (competent-to-incompetent), or three consecutive incompetent behaviors, followed immediately by two competent behaviors (incompetent-to-competent). Morality individuals consisted of faces paired with either three consecutive moral behaviors, followed immediately by two immoral behaviors (moral-to-immoral), or three consecutive immoral behaviors, followed immediately by two moral behaviors (immoral-to-moral).

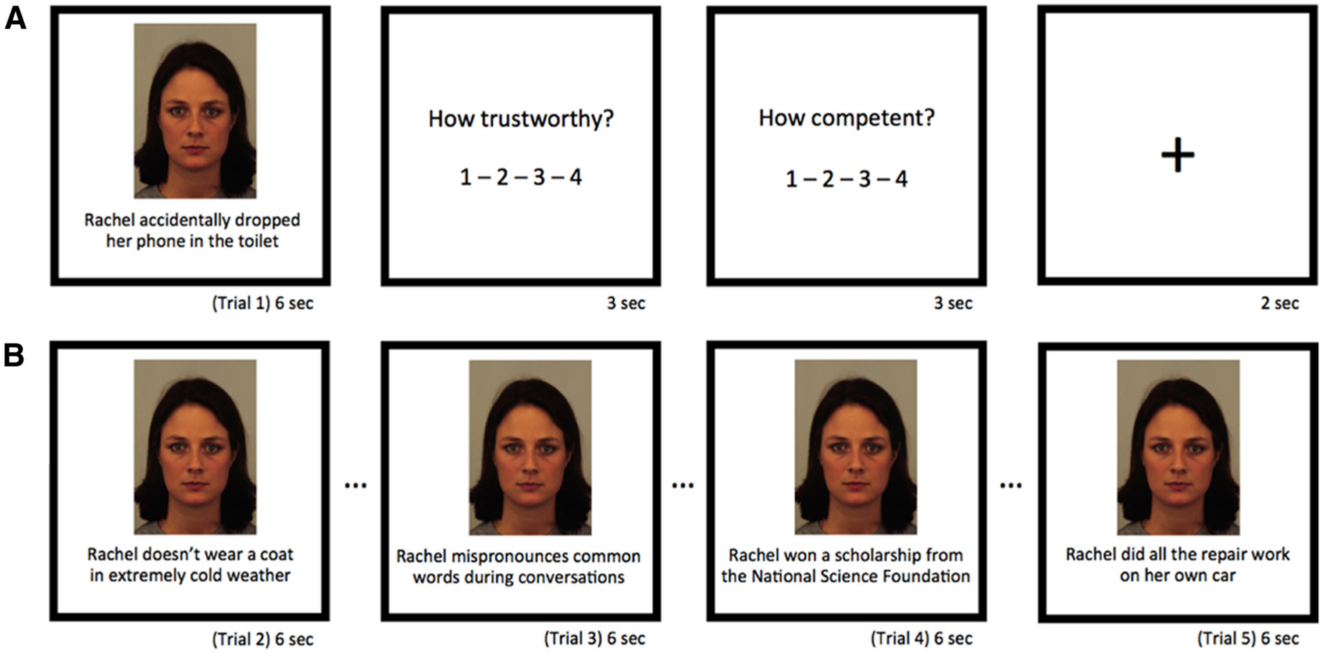
A
Rachel accidentally dropped her phone in the toilet
(Trial 1) 6 sec
How trustworthy?
1 – 2 – 3 – 4
3 sec
How competent?
1 – 2 – 3 – 4
3 sec
+
2 sec
B
Rachel doesn't wear a coat in extremely cold weather
(Trial 2) 6 sec
...
Rachel mispronounces common words during conversations
(Trial 3) 6 sec
...
Rachel won a scholarship from the National Science Foundation
(Trial 4) 6 sec
...
Rachel did all the repair work on her own car
(Trial 5) 6 sec

Figure 1. Example stimuli. ***A***, Example Trial 1. After each face/behavior pair, participants rated individuals on trustworthiness and competence (order counterbalanced between participants). ***B***, Representative behaviors across Trials 2–5, which were viewed consecutively. Note that the last two behaviors are inconsistent with the first three behaviors. Ellipses denote ratings and fixation cross.

For analyses focused on diagnosticity, we collapsed across incompetent-to-competent and moral-to-immoral conditions ("diagnostic updates"), and competent-to-incompetent and immoral-to-moral conditions ("nondiagnostic updates"), respectively.

These distinctions are supported by behavioral ratings recorded from participants (see below) and are consistent with the behavioral asymmetry described in previous research (Skowronski and Carlston, 1987; Wojciszke et al., 1993). Furthermore, for valence analyses, we combined the incompetent-to-competent and immoral-to-moral conditions ("positive updates"), and competent-to-incompetent and moral-to-immoral conditions ("negative updates"), respectively.

Finally, we used a baseline, control condition that accounted for the presence of text on screen paired with each face. Relative to this baseline, we could estimate the change in brain responses for the updating conditions. Control individuals were faces paired with a brief sentence indicating the individual's name (i.e., "This man's name is Ron"). In total, participants encountered 50 individuals—10 corresponding to each of these five conditions.

Behaviors were combined in groups of five, such that each group within a given condition would be approximately equated on both kindness and intelligence, and were presented consistently across all participants. These groups were predetermined before the analysis of behavioral data from Study 1, and as such were not biased by the results of Study 1. Moreover, we determined that the distribution of behaviors among the groups was unbiased, with respect to ratings of kindness and intelligence. Immoral behaviors selected for immoral-to-moral groups were not significantly different from those selected for moral-to-immoral groups, in either kindness ($p = 0.744$) or intelligence ($p = 0.749$). Moral behaviors selected for immoral-to-moral groups were not significantly different from those selected for moral-to-immoral groups, in either kindness ($p = 0.641$) or intelligence ($p = 0.600$). Incompetent behaviors selected for incompetent-to-competent groups were not significantly different from those selected for competent-to-incompetent groups, in either intelligence ($p = 0.318$) or kindness ($p = 0.512$). Finally, competent behaviors selected for incompetent-to-competent groups were not significantly different from those selected for competent-to-incompetent groups, in either intelligence ($p = 0.338$) or kindness ($p = 0.549$).

We counterbalanced faces and behavior groups between participants, such that each face was paired with each type of behavior group an equal number of times. Finally, we created a unique, optimized ordering for each participant, based upon a genetic algorithm (http://wagerlab.colorado.edu/wiki/doku.php/help/ga/genetic_algorithm_for_fmri; Wager and Nichols, 2003) to maximize statistical power.

Procedure. Participants were informed that they would be participating in a study on impression formation. They were told that they would be seeing a series of faces paired with behaviors, and that they would see multiple behaviors paired consecutively with each face. Participants were asked to form an impression of each person, and were informed that some information might run contrary to the impression they had formed so far. We explained that participants might alter their impressions based on new information they learned as the task progressed, and that picturing individuals performing behaviors might aid in forming impressions. Finally, we gave participants explicit instructions as to how to approach the ratings of competence and trustworthiness they would be giving following each behavior. On any given trial, we asked that participants approach each individual rating as an overall impression of a given individual, taking into account everything that had been learned about that person up to that point, and not simply to rate the behavior itself, detached from the context of the other information they had learned about the individual in question.

Participants practiced one full run of the task outside the scanner, so that they could adjust to the timing and design of the task. They were presented with five individuals—comprising faces and behaviors not used in the scanner portion of the task. After reminding participants how they should approach the ratings (i.e., as global impressions of everything they had learned so far about each individual), participants entered the scanner.

In the scanner, participants completed 10 runs. Every run comprised five individuals, one of each condition, each paired with five separate behaviors. Each run began with a 15 s presentation of a fixation cross. Each sequence consisted of five face/behavior presentations. Faces and behaviors were presented together for 6 s. After each presentation of a face/behavior pair, competence and trustworthiness rating slides appeared for 3 s each. (Rating order was counterbalanced between participants, such that half always rated competence first and half always rated trustworthiness first.) Multiple analyses of both behavioral and neural data suggested that there were no consistent differences between participants who rated trustworthiness first and participants who rated competent first.) Subsequently, a fixation cross appeared for 2 s.

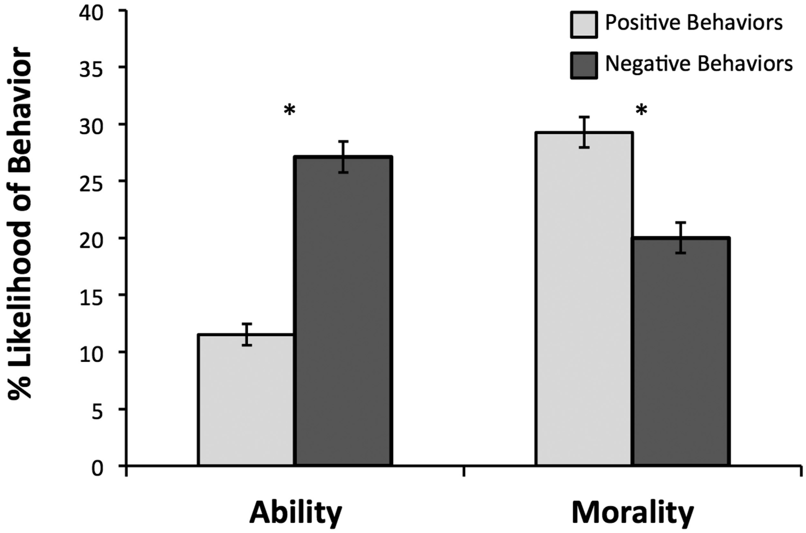

Figure 2. Diagnosticity is an emergent property of perceived frequency. We observed an interaction between valence and domain, such that positive ability behaviors were perceived as less frequent than negative ability behaviors, while negative morality behaviors were perceived as less frequent than positive morality behaviors.

Imaging acquisition. A blood oxygenation level-dependent (BOLD) signal was used as a measure of neural activation. Images using echoplanar imaging (EPI) were acquired using a Siemens 3.0 tesla Allegra head-dedicated scanner with a standard "bird-cage" head coil (TR, 2000 ms; TE, 30 ms; flip angle, 80°; matrix size, 64 × 64). By using 32 interleaved 3 mm axial slices, we were able to achieve nearly whole-brain coverage. Before the primary data acquisition scan, a high-resolution anatomical image (T1-MPRAGE; TR, 2500 ms; TE, 4.3 ms; flip angle, 8°; matrix size, 256 × 256) was acquired for subsequent registration of functional activity to the participant's anatomy and for spatially normalizing data across participants.

Imaging analysis. All fMRI data were analyzed with Analysis of Functional NeuroImages software (Cox, 1996). We discarded the first four EPI images from each run to allow the MR signal to reach steady-state equilibrium. Participants' motion was corrected using a six-parameter 3D motion-correction algorithm following slice scan-time correction. Transient spikes were removed from the signal using the AFNI (Analysis of Functional NeuroImages) program 3dDespike. Subsequently, data were low-pass filtered with a frequency cutoff of 0.1 Hz following spatial smoothing with a 6 mm full-width at half-maximum Gaussian kernel. The signal was then normalized to the percentage signal change from the mean.

Subsequently, we performed a whole-brain analysis testing the main effect of updating [contrasting the last two trials (L2) against the first three trials (F3)], collapsed across ability and morality. Next, we performed whole-brain analyses testing the main effects of domain and valence, and the interaction thereof, during F3 trials (i.e., regardless of updating). Finally, we performed whole-brain analyses testing (1) the interaction between trial order (L2 vs F3) and domain (ability vs morality), (2) the interaction between trial order (L2 vs F3) and valence (positive vs negative), and (3) the interaction between trial order (L2 vs F3) and diagnosticity (diagnostic vs nondiagnostic). For each of these interaction analyses, we also tested the simple effects split by domain (for the diagnosticity and valence interactions) or valence (the domain interaction). These data are reported at a voxelwise threshold of $p < 0.005$. Furthermore, to select a minimum cluster size for corrected significance ($p < 0.05$), we performed a Monte Carlo simulation of null-hypothesis data, using the AlphaSim program included in the AFNI package. The Monte Carlo simulation indicated that a minimum cluster size of 31 voxels was appropriate.

To generate parameter estimates, we performed voxelwise multiple regression on each participant's preprocessed imaging data. Twenty-five regressors of interest (five 6000 ms trials per individual × five conditions) were convolved with a canonical hemodynamic response function and entered into our general linear model (GLM). Additionally, we included several regressors of no interest, including head motion estimates and time points representing rating slide presentations. Each participant's parameter estimate maps were projected into Talairach space (Talairach and Tournoux, 1988) before performing any group-level analyses.

We performed a second, separate GLM in which we modeled neural responses as a function of (1) update magnitude and (2) perceived frequency of behaviors. The first parametric regressor tracked the absolute change in ratings (in the relevant domain) from the present trial to the previous trial, for each individual participant. For instance, if an individual varying on ability was given a competence rating of 4 on Trial 3 and a rating of 1 on Trial 4, the update magnitude regressor would reflect this change with a value of 3 on Trial 4. (As learning on Trial 1 did not constitute an update, but rather the formation of an initial impression, the update magnitude regressor always had a value of 0 at Trial 1.) To construct the second parametric regressor, we assigned a value to each face/behavior pair related to the consensus of perceived frequency of that behavior, based on the data collected in Study 1. Specifically, since the frequency values collected in Study 1 were out of 100, the regressor values were reverse scored, so that rare behaviors would be associated with increased activity.

Conjunction analysis. We performed two conjunction analyses designed to test for overlapping activity shared between (1) the L2 versus F3 contrast and the parametric analysis focused on update magnitude, and (2) the diagnosticity interaction contrast and the parametric analysis focused on perceived behavioral frequency. To do so, thresholded brain maps (voxelwise thresholding, $p < 0.005$; corrected as explained above) were first converted to binary maps, which were then assigned specific values, such that different colors would be associated with different maps in the resulting conjunction map. Specifically, we assigned a value of 1 to all clusters from both contrast maps and a value of 2 to all clusters from both parametric analyses. We next added the appropriate maps together using AFNI command line tools, such that resulting areas of overlap received a value of 3.

Results

Study 1: perceived frequency guides behavioral diagnosticity

A 2 × 2 ANOVA revealed a significant interaction between domain (ability vs morality) and valence (negative vs positive; $F_{(1,79)} = 321.11$, $p < 0.001$, $\eta_p^2 = 0.80$). Specifically, participants indicated that competent behaviors occur less frequently in the environment than incompetent behaviors ($t_{(79)} = 14.17$, $p < 0.0001$), while immoral behaviors occur less frequently than moral behaviors ($t_{(79)} = 6.76$, $p < 0.0001$; Fig. 2). These findings are consistent with the hypothesis that the perceived frequency of behaviors underlies the asymmetric updating in the morality and ability domains.

Study 2: behavioral diagnosticity influences neural responses during updating

Behavioral analyses

We computed separate averages of both trustworthiness and competence ratings across F3 and L2 behaviors, isolating participants' evaluations of our targets before and after the introduction of inconsistent information.

The four-way interaction among domain (ability vs morality), valence (positive-to-negative vs negative-to-positive), order (first three trials vs last two trials), and rating (trustworthiness vs compe-

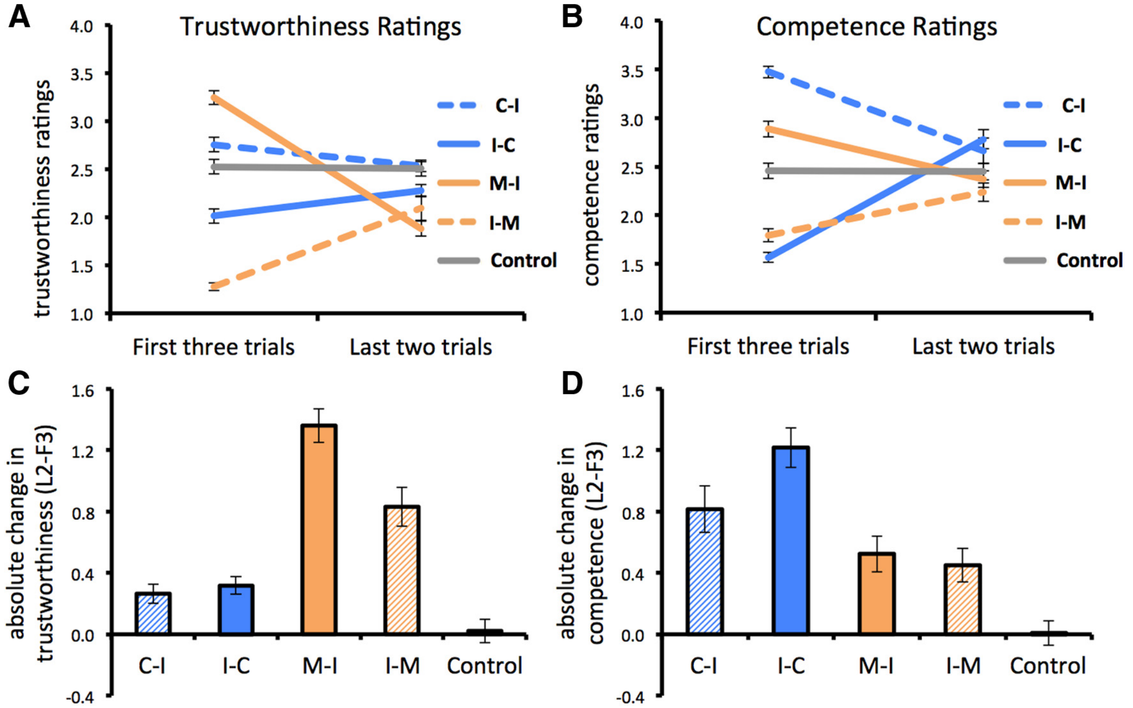

Figure 3. Behavioral evidence of impression updating. ***A***, ***B***, Participants' ratings of trustworthiness changed more when evaluating individuals varying on morality (***A***), while their ratings of competence changed more when evaluating individuals varying on ability (***B***). Moreover, we observed an asymmetry between the two domains in terms of which information had the strongest impact on impression updating. ***C***, Negative information was more diagnostic in the morality domain, as the absolute change in trustworthiness ratings was largest for the moral-to-immoral individuals. The absolute change in trustworthiness was larger for individuals varying on morality (orange bars) than those varying on ability (blue bars). ***D***, Positive information was more diagnostic in the ability domain, as the absolute change in competence ratings was largest for the incompetent-to-competent individuals. The absolute change in competence was larger for individuals varying on ability (blue bars) than those varying on morality (orange bars). C-I, Competent-to-incompetent; I-C, incompetent-to-competent; M-I; moral-to-immoral; I-M, immoral-to-moral.

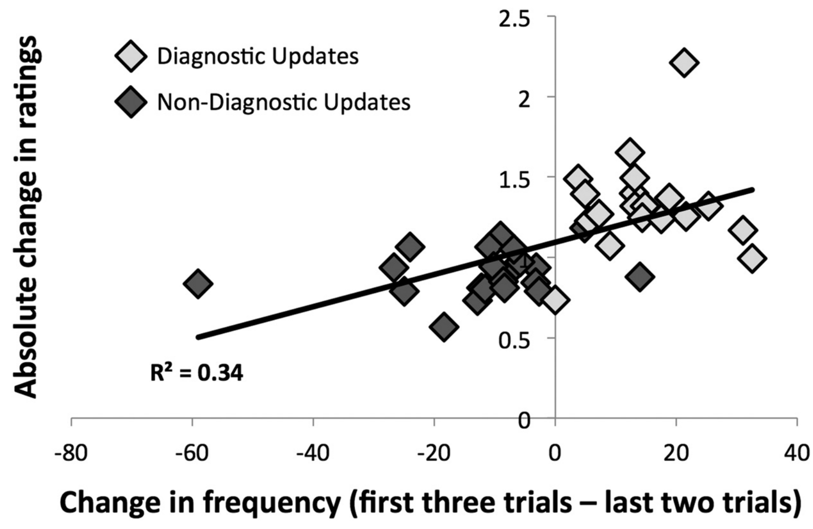

Figure 4. Correlation between update magnitude and changes in perceived frequency. Larger update magnitudes (i.e., absolute changes in relevant ratings) were positively associated with larger changes in the perceived frequency of behaviors (collected externally). As behaviors were grouped in a predetermined fashion, each marker represents one individual, comprising five specific behaviors. Positive values on the x-axis denote individuals for whom the first three behaviors were perceived as more frequent, on average, than the last two. Negative values denote individuals for whom the first three behaviors were perceived as less frequent, on average, than the last two. Dark gray markers represent individuals containing nondiagnostic updates (i.e., a shift from immoral to moral or competent to incompetent behaviors), while light gray markers represent individuals containing diagnostic updates (i.e., a shift from moral to immoral or incompetent to competent behaviors).

tence) was significant ($F_{(1,22)} = 131.18, p < 0.0001, \eta^2_p = 0.86$), reflecting the predicted effects. First, within each type of rating, the three-way interactions among domain, valence, and order were significant (Fig. 3*A*,*B*; trustworthiness: $F_{(1,22)} = 120.62$, $p < 0.0001$, $\eta^2_p = 0.85$; competence: $F_{(1,22)} = 71.02, p < 0.0001, \eta^2_p = 0.76$). These interactions indicated that whereas trustworthiness ratings changed more for morality than competence individuals (valence × order: $F_{(1,22)} = 88.84, p < 0.0001, \eta^2_p = 0.80$ vs $F_{(1,22)} = 15.91$, $p = 0.0001$, $\eta^2_p = 0.42$, respectively), competence ratings changed more for competence than morality individuals ($F_{(1,22)} = 56.52$, $p < 0.0001$, $\eta^2_p = 0.72$ vs $F_{(1,22)} = 18.69$, $p < 0.0001$, $\eta^2_p = 0.46$, respectively).

These effects are clearly seen in the absolute changes in trustworthiness and competence ratings (Fig. 3*C*,*D*). For trustworthiness ratings, the absolute change was larger for morality individuals than for ability individuals (mean, 0.81; SE, 0.08; $p < 0.0001$). Moreover, the absolute change was larger for moral-to-immoral than immoral-to-moral individuals ($t_{(22)} = 7.07$, $p < 0.0001$), but did not differ significantly between competent-to-incompetent and incompetent-to-competent individuals ($t_{(22)} = 1.11$, $p = 0.28$). In contrast, for competence ratings, the absolute change was larger for ability than for morality individuals (mean, 0.53; SE, 0.06; $p < 0.0001$). Further, the absolute change was larger for incompetent-to-competent than competent-to-incompetent individuals ($t_{(22)} = 4.65$, $p < 0.0001$), but did not differ significantly between moral-to-

Table 1. Regions displaying a main effect of updating: differential responses to first three and last two trials

Region	Hemi	Voxels	*x*	*y*	*z*
Last two trials $>$ first three trials					
Activity spanning dorsomedial, dorsolateral, ventrolateral, and rostrolateral prefrontal cortex*	R	2129	31.5	58.5	−3.5
Superior temporal sulcus/anterior temporal lobe**	R	395	67.5	−34.5	−6.5
Inferior parietal lobule/temporoparietal junction	R	344	46.5	−61.5	44.5
Inferior parietal lobule/temporoparietal junction	L	138	−46.5	−58.5	41.5
Rostrolateral prefrontal cortex/inferior frontal gyrus	L	132	−28.5	58.5	−3.5
Precuneus	—	111	1.5	−49.5	23.5
Dorsolateral prefrontal cortex	L	111	−43.5	19.5	41.5
Posterior cingulate cortex	—	54	1.5	−25.5	26.5
Ventrolateral prefrontal cortex	L	53	−28.5	19.5	−3.5
First three trials $>$ last two trials					
Inferotemporal cortex/primary visual cortex	R/L	3901	40.5	−55.5	−18.5
Posterior thalamus	R	48	4.5	−31.5	−0.5
Mid-cingulate cortex	—	43	1.5	1.5	38.5
Precentral gyrus	R	41	40.5	1.5	29.5

Group results ($N = 23$), $p < 0.05$ FDR corrected; voxelwise threshold, $p < 0.005$; minimum cluster-size threshold, 16 voxels. Coordinates refer to the peak voxel in Talairach space. For each cluster, we report its hemisphere (Hemi) and size in voxels (Voxels). *This large activation cluster contained peaks (coordinates are peak voxel in Talairach space, *x*, *y*, *z*) in dorsomedial (7.5, 61.5, 26.5), dorsolateral (43.5, 16.5, 44.5), ventrolateral (49.5, 19.5, −3.5), and rostrolateral PFC (31.5, 58.5, −3.5). **This cluster contained peaks in superior temporal sulcus (67.5, −34.5, −6.5) and anterior temporal lobe (46.5, 19.5, −30.5). R, Right; L, left.

Table 2. Regions displaying effects of domain and valence, first three trials only

Region	Hemi	Voxels	*x*	*y*	*z*
Main effect of domain					
Posterior superior temporal sulcus	R	62	46.5	−76.5	−3.5
Main effect of valence					
Inferotemporal cortex	R	89	52.5	−58.5	−15.5
Cuneus	R	41	13.5	−88.5	20.5
Precuneus		36	1.5	−67.5	23.5
Inferior frontal gyrus	L	32	−37.5	40.5	23.5
Interaction between domain and valence					
Inferior frontal gyrus	L	237	−49.5	28.5	23.5
Superior temporal sulcus	L	116	−52.5	−40.5	−12.5
Ventrolateral prefrontal cortex	L	54	−43.5	28.5	−9.5
Fusiform gyrus	L	38	−40.5	−73.5	−12.5
Precentral gyrus	L	34	−13.5	−28.5	68.5
Medial prefrontal cortex		31	−4.5	64.5	8.5
Anterior temporal lobe	L	28*	−55.5	1.5	−3.5
Inferotemporal cortex	L	21*	−55.5	−55.5	−18.5
Ventrolateral prefrontal cortex	R	20*	46.5	25.5	−9.5

Group results ($N = 23$), $p < 0.05$ FDR corrected; voxelwise threshold, $p < 0.005$; minimum cluster-size threshold, 31 voxels. Coordinates refer to the peak voxel in Talairach space. For each cluster, we report its hemisphere (Hemi) and size in voxels (Voxels). R, Right; L, left.

immoral and immoral-to-moral individuals ($t_{(22)} = 1.18$, $p = 0.25$).

We also examined the relationship between the magnitude of updating and perceptions of frequency. To do so, we computed the absolute change in ratings between L2 and F3 trials for relevant ratings (i.e., competence ratings for individuals varying in the ability domain and trustworthiness ratings for individuals varying in the morality domain), and tested the Pearson correlation with the average change in perceived frequency (F3 − L2), represented by ratings obtained in Study 1. (We note that, while the behaviors from Study 1 were used in Study 2, behaviors were arranged into predetermined groups for use in Study 2 before the analysis of behavioral data from Study 1, so as not to bias the design of Study 2.) Increases in participants' update magnitudes correlated with increases in changes in perceived frequency ($r = 0.58$, $p < 0.0001$; Fig. 4).

These results indicate that (1) trustworthiness and competence ratings were sensitive to their respective domains of morality and ability; (2) whereas negative information had a stronger influence on trustworthiness ratings of individuals varying on morality, positive information had a stronger influence on competence ratings of individuals varying on ability; and (3) there is a direct relationship between the magnitude of updating and consensus perceptions of statistical frequency.

Neuroimaging analyses

Main effect of updating. A whole-brain analysis testing the main effect of updating (L2 $>$ F3, p(corrected) < 0.05) revealed a set of regions showing an enhanced BOLD response during L2 trials, compared with F3 trials, including dorsomedial prefrontal cortex (dmPFC); inferior frontal gyrus (IFG), extending down through bilateral rostrolateral PFC; bilateral ventrolateral PFC (vlPFC); right superior temporal sulcus (STS); inferior parietal lobule (IPL), including the temporoparietal junction (TPJ); precuneus; and posterior cingulate cortex (Table 1). (Regions showing an enhanced BOLD response during F3 trials are also detailed in Table 1.)

Main effect of domain, F3 trials only. During F3 trials, we observed preferential responses to morality behaviors (compared with ability behaviors) in a region of right posterior STS extending into the TPJ (Table 2, Main effect of domain). No regions showed preferential responses to ability behaviors during F3 trials.

Main effect of valence, F3 trials only. During F3 trials, we observed preferential responses to positively valenced behaviors, compared with negatively valenced behaviors, in left dorsolateral prefrontal cortex (dlPFC), right inferotemporal cortex, cuneus, and precuneus (Table 2, Main effect of valence). No regions showed preferential responses to negatively valenced behaviors during F3 trials.

Interaction between domain and valence, F3 trials only. During F3 trials, we observed preferential responses to diagnostic behaviors (i.e., competent and immoral), compared with nondiagnostic behaviors (i.e., incompetent and moral), across a large set of regions, including left IFG, left STS, left vlPFC, left fusiform gyrus (FG), and mPFC (Table 2, Interaction between domain and valence). We also observed similar activity in left anterior temporal lobe (ATL) and right vlPFC, though these regions did not surpass cluster thresholding ($k = 31$). No regions showed preferential responses to nondiagnostic behaviors during F3 trials.

Interaction analysis: domain-specific updating. The domain interaction, testing for updating-specific activity that was preferential for either the ability or morality domain, revealed no domain-specific increases in activity associated with updating.

Interaction analysis: valence-specific updating. Further, we tested whether negative information in either domain might have a stronger influence than positive information. While this valence interaction (Table 3) failed to yield any regions displaying activity specific to updating based on negative information, we observed several regions showing the opposite effect. Medial orbitofrontal cortex (mOFC), precuneus, and cuneus activity was stronger during L2 trials when new, inconsistent information was positive—i.e., implied high morality or high competence—compared with negative.

We extracted parameter estimates from each of these regions to assess whether this pattern of activity was consistent across domains. Separate ANOVAs indicated that the three-way interactions among domain, valence, and order were nonsignificant in mOFC, cuneus, and precuneus, suggesting that the response to positively valenced behaviors presented on L2 trials did not vary between the ability and morality domains (Table 3).

Interaction analysis: diagnosticity-specific updating. The diagnosticity interaction contrast revealed activity in bilateral vlPFC,

Table 3. Regions displaying an interaction between trial order (first three vs last two trials) and behavioral valence

Region	Hemi	Voxels	*x*	*y*	*z*	Ability domain	C-I	I-C	Morality domain	M-I	I-M
Medial orbitofrontal cortex		187	−1.5	31.5	−0.5	Yes*	−0.548**	0.573**	Yes*	−0.586***	0.621**
Precuneus		98	−1.5	−64.5	23.5	Yes***	0.297†	0.765**	Yes*	−0.457‡	0.785*
Cuneus	R	45	10.5	−91.5	17.5	Yes**	−1.04*	−0.182‡	Yes*	−1.36*	−0.346‡

Group results ($N = 23$), $p < 0.05$ FDR corrected; voxelwise threshold, $p < 0.005$; minimum cluster-size threshold, 31 voxels. Extracting parameter estimates confirmed that activity in all three regions increased when participants were updating based on positive information only (i.e., competent behaviors or moral behaviors). Coordinates refer to the peak voxel in Talairach space (x, y, z). For each cluster, we report its hemisphere (Hemi) and size in voxels (Voxels). Ability domain and Morality domain indicate whether valence effects were significant in either domain. We also detail the change in parameter estimates from F3 to L2 trials for competent-to-incompetent (C-I), incompetent-to-competent (I-C), moral-to-immoral (M-I) and immoral-to-moral (I-M) individuals. *$p < 0.001$; **$p < 0.01$; ***$p < 0.05$; †NS ($p > 0.10$); ‡$p < 0.10$.

Table 4. Regions displaying an interaction between trial order (first three versus last two trials) and behavioral diagnosticity

Region	Hemi	Voxels	*x*	*y*	*z*	Ability domain	C-I	I-C	Morality domain	M-I	I-M
Ventrolateral prefrontal cortex/inferior frontal gyrus	L	303	−49.5	25.5	−6.5	Not	−0.219†	0.053†	Yes*	−0.603*	0.953*
Ventrolateral prefrontal cortex	R	69	52.5	22.5	−3.5	Yes**	−0.177†	0.677*	Yes**	0.959***	0.064†
Superior temporal sulcus	L	49	−61.5	−46.5	5.5	Not	−0.102†	0.116†	Yes*	−0.377*	1.036*
Fusiform gyrus	L	36	−40.5	−67.5	−12.5	Yes***	−1.323*	−0.215†	Not	−0.462‡	−0.719***

Group results ($N = 23$), $p < 0.05$ FDR corrected; voxelwise threshold, $p < 0.005$; minimum cluster-size threshold, 31 voxels. Extracting parameter estimates confirmed that activity in all three regions increased when participants were updating based on diagnostic information only (i.e., competent behaviors or immoral behaviors). Coordinates refer to the peak voxel in Talairach space (x, y, z). For each cluster, we report its hemisphere (Hemi) and size in voxels (Voxels). Ability domain and Morality domain indicate whether diagnosticity effects were significant in either domain. We also detail the change in parameter estimates from F3 to L2 trials for competent-to-incompetent (C-I), incompetent-to-competent (I-C), moral-to-immoral (M-I) and immoral-to-moral (I-M) individuals. R, Right; L, left. *$p < 0.001$; **$p < 0.05$; ***$p < 0.01$; †NS ($p > 0.10$); ‡$p < 0.10$.

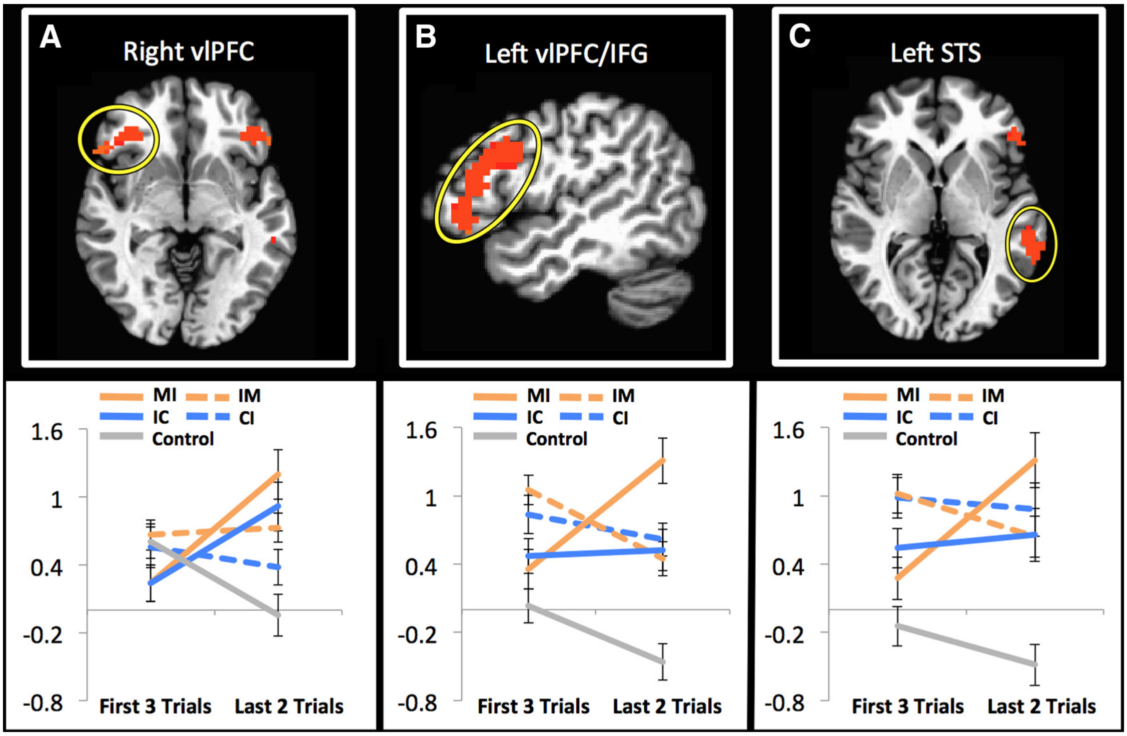

Figure 5. The impact of behavioral diagnosticity on neural responses associated with impression updating [$N = 23$, $p < 0.05$ false discovery rate (FDR) corrected]. ***A–C***, Right vlPFC (***A***), left vlPFC/IFG (***B***), and left STS (***C***) all show increased activity when updating impressions based upon diagnostic information. Expanded analysis of parameter estimates suggested that while activity in right vlPFC was associated with diagnostic updating in either domain, left vlPFC/IFG and left STS activity was driven primarily by updates regarding immoral behaviors. (In the control condition, participants saw faces paired with names only.)

left IFG, left STS, and left FG, which was stronger during L2 trials when new, inconsistent information was diagnostically valuable—i.e., implied high immorality or high competence—compared with nondiagnostic (Table 4; Fig. 5).

We extracted parameter estimates from each of these regions and ran separate ANOVAs at each level of domain to assess whether the pattern of diagnosticity-specificity was consistent across domains. These analyses indicated that, while right vlPFC activity increased when diagnostic behaviors from either the ability and morality domains appeared on L2 trials, activity in left vlPFC/IFG and left STS was primarily associated with updating in the morality domain, while activity in left fusiform face area was primarily associated with updating in the ability domain (Table 4).

Parametric analyses. To further interrogate the neuroimaging data, we performed separate whole-brain analyses aimed at identifying brain regions whose activity covaried parametrically with (1) the absolute magnitude of updates from trial-to-trial and (2) perceptions of behavioral frequency. (See Materials and Methods for a full description of how these regressors were assembled.)

These two parametric analyses implicated two distinct sets of regions. On the one hand, we observed activity covarying with update magnitude in dmPFC, bilateral IPL/TPJ, right ATL, right STS, right dlPFC, and right anterior insula/vlPFC (Table 5, Ac-

Table 5. Regions displaying parametric effects of update magnitude and behavioral frequency

Region	Hemi	Voxels	x	y	z
Activity covarying parametrically with update magnitude					
Dorsomedial prefrontal cortex	R	400	10.5	10.5	62.5
Ventrolateral prefrontal cortex	R	144	37.5	25.5	−15.5
Inferior parietal lobule/temporoparietal junction	R	105	55.5	−55.5	32.5
Anterior temporal lobe	R	80	40.5	22.5	−30.5
Superior temporal sulcus	R	49	67.5	−28.5	−3.5
Inferior parietal lobule/temporoparietal junction	L	42	−49.5	−61.5	26.5
Dorsolateral prefrontal cortex	R	36	43.5	13.5	44.5
Activity covarying parametrically with behavioral frequency					
Parahippocampal gyrus	L	90	−55.5	−31.5	−15.5
Lingual gyrus	L	84	34.5	−82.5	−15.5
Ventrolateral prefrontal cortex	R	55	−46.5	31.5	−0.5
Superior temporal sulcus	L	31	−61.5	−49.5	2.5

Group results ($N = 23$), $p < 0.05$ FDR corrected; voxelwise threshold, $p < 0.005$; minimum cluster-size threshold, 16 voxels. Coordinates refer to the peak voxel in Talairach space (*x*, *y*, *z*). For each cluster, we report its hemisphere (Hemi) and size in voxels (Voxels). R, Right; L, left.

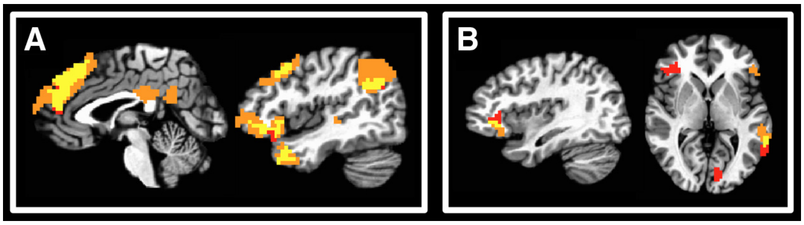

Figure 6. Conjunction maps of contrasts and parametric analyses. ***A***, We observed significant overlap (yellow) in dmPFC, bilateral TPJ/IPL, right ATL, right dlPFC, and right vlPFC between activity associated with the last two trials (orange) and activity covarying parametrically with update magnitude (red). ***B***, Right vlPFC and left STS displayed significant overlap (yellow) between activity tracking diagnosticity-related updating (orange) and activity covarying parametrically with consensus perceptions of behavioral frequency (red).

tivity covarying parametrically with update magnitude). On the other hand, we observed activity covarying with perceived behavioral frequency in right vlPFC, left posterior STS (pSTS), left parahippocampal gyrus, and left lingual gyrus (Table 5, Activity covarying parametrically with behavioral frequency).

Noting the theoretical similarities between (1) the L2 > F3 contrast and regions whose activity covaried parametrically with update magnitude, and (2) the diagnosticity-specific updating contrast and regions whose activity covaried parametrically with perceived behavioral frequency, we performed conjunction analyses aimed at identifying overlap between these parametric analyses and contrasts. We reasoned that these two different types of analyses could potentially yield convergent evidence regarding brain regions involved in updating impressions and assessing frequency-based behavioral diagnosticity, respectively. Significant overlap is shown in Figure 6, with contrast maps displayed in orange, parametric maps displayed in red, and overlap displayed in yellow.

In the former conjunction analysis, we observed almost complete correspondence between regions preferentially active during the last two trials and regions displaying parametric covariation with update magnitude, with significant overlap in dmPFC, bilateral TPJ/IPL, right ATL, right dlPFC, and right vlPFC (Fig. 6*A*). In the latter conjunction analysis, we observed overlap between regions preferentially active when updating based on diagnostic behaviors and regions displaying parametric covariation with perceived behavioral frequency in right vlPFC and left STS (Fig. 6*B*).

Discussion

Across two studies, we examined the behavioral and neural impacts of behavioral diagnosticity on impression updating. We demonstrate that the diagnosticity of a behavior is directly related to its perceived frequency in real life, and, moreover, that updating based on rare, diagnostic behaviors preferentially evokes activity in right vlPFC. Furthermore, updating based specifically on immoral behaviors evokes activity in left vlPFC, left IFG, and left STS.

The Study 1 results mirror the prevailing conceptualization of diagnosticity within the ability and morality domains (Skowronski and Carlston, 1987; Wojciszke, 2005), wherein highly competent and highly immoral actions are more diagnostic than their counterparts. By extension, these results suggest that diagnosticity is an emergent property of perceived frequency. On some level, this is perhaps not surprising—judgments of competence and morality are not made based on absolute, objective metrics, but in relation to the behaviors of other individuals. If an individual is judged to be extremely competent, it is a reflection of the fact that, on average, other people behave less competently than the individual in question. Similarly, varying perceptions of the frequency of behaviors may reflect real-world differences in experience. Moreover, these perceptions may be malleable via either bottom-up or top-down manipulation. Future work should explore how these perceptions of behavioral frequency form, as well as the degree to which they can be altered.

The behavioral results of Study 2 are also consistent with a diagnostic asymmetry between the ability and morality domains. Participants reported larger changes in competence evaluations in response to competent behaviors presented on L2 trials, compared with incompetent behaviors, while morality evaluations were more strongly impacted by immoral behaviors presented on L2 trials than moral behaviors. More importantly, we observed a strong correlation between the degree to which participants updated their impressions and the perceived frequency of the behaviors provoking those updates, providing direct evidence for the connection between diagnosticity and frequency suggested by the results of Study 1. Specifically, participants updated their impressions more strongly when new, inconsistent information was also perceived as rare, compared with the preceding information.

Our initial neuroimaging analysis contrasting activity during L2 trials against activity during F3 trials was consistent with previous investigations of impression updating (Cloutier et al., 2011b; Ma et al., 2012; Mende-Siedlecki et al., 2013). These results replicated a network of regions involved in updating impressions of others. In addition, we observed that activity in several members of this network (dmPFC, right STS, bilateral IPL/TPJ) covaried parametrically with the magnitude of participants' updates from trial to trial, further confirming that these regions play a critical role in the updating process. While it is possible to hypothesize regarding the individual contributions of these regions based on prior research, the design of the present study renders this a speculative prospect. Nevertheless, the interaction analyses and parametric analyses do offer more concrete evidence.

Our interaction analyses revealed a set of regions (bilateral vlPFC, left IFG, and left STS) that were strongly influenced by the diagnostic value of behaviors, rather than their domain or valence. Critically though, examining the parameter estimates extracted from these diagnosticity-sensitive ROIs suggested that only the right vlPFC was equally influenced by diagnostic behaviors in both the ability and morality domains. Activity in the left vlPFC/IFG and left STS was primarily associated with updates elicited by immoral behaviors, in particular. However, activity in a similar region of left STS, as well as right vlPFC, covaried parametrically with consensus perceptions of behavioral frequency.

These analyses provide strong, convergent evidence suggesting that the right vlPFC plays a key role in updating based on low-frequency, diagnostic behavioral information (regardless of domain). The left STS is similarly recruited by low-frequency, highly diagnostic behaviors, though we concede that its role in updating is less clear cut than that of the right vlPFC. More importantly, we show a neural link between diagnosticity-specific activity and activity covarying with perceived behavioral frequency, in direct parallel to the behavioral links between diagnosticity and frequency demonstrated in Study 1.

Previous work suggests that anterior vlPFC (IFG pars orbitalis) is responsible for controlled, top-down retrieval of stored conceptual representations, while recruitment of a more posterior region of mid-vlPFC (IFG pars triangularis) reflects the resolution of competition between accessed representations following retrieval (Badre and Wagner, 2005, 2007; Souza et al., 2009; Satpute et al., 2013). While this pattern of functionality is typically left lateralized, some researchers have suggested parallels between left and right vlPFC function (Badre and Wagner, 2007; Kuhl and Wagner, 2009), especially with regard to the operations of the mid-vlPFC under conditions of decision uncertainty (Levy and Wagner, 2011). Ultimately, updating a person impression should follow similar logic—an existing impression is accessed and compared against incoming information, and, if the new information is diagnostically valuable, the impression is updated. Moreover, we note highly relevant recent work implicating this region in the integration of new behavioral information with initial impressions based on facial appearance (referred to as lateral orbitofrontal cortex, but anatomically consistent; Kim et al., 2012; Bhanji and Beer, 2013).

We also observed similar patterns of diagnosticity-related activity in the STS. Increased activity in STS during diagnostic updates may reflect a signal akin to a social prediction error, as it has been argued that STS activity increases when social agents violate prior expectancies (Behrens et al., 2008; Frith and Frith, 2010). In the context of the present research, this signal should be most salient when that inconsistency is most diagnostic. While the precise computational contributions of the vlPFC and STS remain unclear, the current results suggest that when learning about other people, the brain may be tracking low-level statistical properties of behavior in service of the superordinate goal of updating impressions.

The neuroimaging results of Study 2 also clarify the role of behavioral valence in impression updating. While some work in social psychology has suggested that affectively negative stimuli are ultimately more powerful than their positive counterparts (Baumeister et al., 2001), we did not observe any updating-related activity that was specific to negative behaviors (i.e., immoral and incompetent behaviors). Instead, we observed a small set of regions that responded preferentially when updating was based on positive behaviors, including the mOFC. Notably, analysis of the extracted parameter estimates suggested that activity in these regions increased when updating based on positively valenced behaviors from either domain; in other words, updating based on both competent and moral behaviors elicited a response in the mOFC, cuneus, and precuneus. Ultimately, this result is consistent with previous work observing mOFC activity associated with moral actions (Zahn et al., 2009; Tsukiura and Cabeza, 2011) and, more generally, with work linking mOFC to reward processing (Rolls, 2000; O'Doherty, 2004), specifically computations regarding subjective value (Noonan et al., 2011; Rushworth et al., 2011).

While some researchers have argued that the ability and morality domains are inherently distinct from one another (Reeder, 1993, 2006), we observed few differences between the neural signatures of impression formation based upon ability and morality information. Only the right pSTS (extending into right TPJ) responded preferentially toward the morality domain, consistent with previous work implicating this region in moral judgment (Young and Saxe, 2009). Moreover, no regions showed preferential updating-related activity toward either domain when collapsed across valence.

The results of any investigation of behavior-based impression formation or impression updating are necessarily bounded by its behavioral stimuli. While we attempted to ensure that the behaviors selected in Studies 1 and 2 were not inherently asymmetric in nature, we cannot be sure that this collection of behaviors represents a typical distribution of behavior that the average person would either experience or perform. Future work should provide empirical evidence for whether or not these asymmetries in behavior truly exist in real life. Moreover, while we have attempted to highlight general trends that guide updating in the ability and morality domains, more research is needed to examine additional motivational influences on the updating process (e.g., outcome dependency; Ames and Fiske, 2013). Finally, the somewhat mixed results of the diagnosticity interaction highlight the need for future work using more sophisticated paradigms. The current design is relatively underpowered, with each category of individual (e.g., incompetent-to-competent) presented only 10 times throughout the entire experiment. A better-powered design, specifically one designed for use in the context of model-based fMRI (Daw et al., 2005, 2011; O'Doherty et al., 2007), would be a marked improvement. Such a design has the potential to advance our understanding beyond a simple catalog of which regions play a role in the updating process, and toward a comprehensive account of the specific computational contributions of each member of the updating network.

Ultimately, these results illuminate a classic finding in the social psychology literature: an asymmetry wherein competent and immoral behaviors impact our impressions of others more profoundly than incompetent and moral behaviors. While the ability and morality domains have been conceptualized as being distinct from one another and operating by different rules, our results indicate that the same statistical principle underlies impression updating in both domains—the informational value of behavior.

References

Ames DL, Fiske ST (2013) Outcome dependency alters the neural substrates of impression formation. Neuroimage 83:599–608. CrossRef Medline

Badre D, Wagner AD (2005) Frontal lobe mechanisms that resolve proactive interference. Cereb Cortex 15:2003–2012. CrossRef Medline

Badre D, Wagner AD (2007) Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45:2883–2901. CrossRef Medline

Baron SG, Gobbini MI, Engell AD, Todorov A (2011) Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. Soc Cogn Affect Neurosci 6:572–581. CrossRef Medline

Baumeister R, Bratslavsky E, Finkenauer C, Vohs K (2001) Bad is stronger than good. Rev Gen Psychol 5:323–370. CrossRef

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245–249. CrossRef Medline

Bhanji JP, Beer JS (2013) Dissociable neural modulation underlying lasting first impressions, changing your mind for the better, and changing it for the worse. J Neurosci 33:9337–9344. CrossRef Medline

Brycz H, Wojciszke B (1992) Personality impressions on ability and morality dimensions. Polish Psychol Bull 23:223–236.

Cloutier J, Kelley WM, Heatherton TF (2011a) The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. Soc Neurosci 6:63–75. CrossRef Medline

Cloutier J, Gabrieli JD, O'Young D, Ambady N (2011b) An fMRI study of violations of social expectations: when people are not who we expect them to be. Neuroimage 57:583–588. CrossRef Medline

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173. CrossRef Medline

Crocker J, Fiske S, Taylor S (1984) Schematic bases of belief change. In: Attitudinal judgment (Eiser J, Ed), pp 197–226. New York: Springer.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711. CrossRef Medline

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215. CrossRef Medline

De Bruin ENM, Van Lange PAM (2000) What people look for in others: influences of the perceiver and the perceived on information selection. Pers Soc Psychol Bull 26:206–219. CrossRef

Fiske ST (1980) Attention and weight in person perception: the impact of negative and extreme behavior. J Pers Soc Psychol 38:889–906. CrossRef

Fiske ST, Cuddy AJ, Glick P (2007) Universal dimensions of social perception: warmth and competence. Trends Cogn Sci 11:77–83. CrossRef Medline

Freeman JB, Schiller D, Rule NO, Ambady N (2010) The neural origins of superficial and individuated judgments about ingroup and outgroup members. Hum Brain Mapp 31:150–159. CrossRef Medline

Frith U, Frith C (2010) The social brain: allowing humans to boldly go where no other species has been. Philos Trans R Soc Lond B Biol Sci 365:165–176. CrossRef Medline

Fuhrman RW, Bodenhausen GV, Lichtenstein M (1989) On the trait implications of social behaviors: kindness, intelligence, goodness, and normality ratings for 400 behavior statements. Behav Res Methods Instrum Comput 21:587–597. CrossRef

Kim H, Choi MJ, Jang IJ (2012) Lateral OFC activity predicts decision bias due to first impressions during ultimatum games. J Cogn Neurosci 24:428–439. CrossRef Medline

Kubicka-Daab J (1989) Positivity and negativity effects in impression formation: differences in processing information about ability and morality dispositions. Polish Psychol Bull 20:295–307.

Kuhl BA, Wagner AD (2009) Strategic control of memory. In: Encyclopedia of neuroscience, Vol 9 (Squire L, Ed), pp 437–444. Oxford: Academic.

Levy BJ, Wagner AD (2011) Cognitive control and right ventrolateral prefrontal cortex: reflexive attention, motor inhibition, and action updating. Ann N Y Acad Sci 1224:40–62. CrossRef Medline

Lewicka M, Czapinski J, Peeters G (1992) Positive-negative asymmetry or "when the heart needs a reason." Eur J Soc Psychol 22:425–434.

Lundqvist D, Flykt A, Öhman A (1998) The Karolinska directed emotional faces—KDEF, CD ROM. Stockholm, Sweden: Department of Clinical Neuroscience, Psychology section, Karolinska Institutet ISBN 91-630-7164-9. Available at: http://www.emotionlab.se/resources/kdef.

Ma N, Vandekerckhove M, Baetens K, Van Overwalle F, Seurinck R, Fias W (2012) Inconsistencies in spontaneous and intentional trait inferences. Soc Cogn Affect Neurosci 7:937–950. CrossRef Medline

Mende-Siedlecki P, Cai Y, Todorov A (2013) The neural dynamics of updating person impressions. Soc Cogn Affect Neurosci 8:623–631. CrossRef Medline

Mitchell JP, Neil Macrae C, Banaji MR (2005) Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. Neuroimage 26:251–257. CrossRef Medline

Mitchell JP, Macrae CN, Banaji MR (2004) Encoding-specific effects of social cognition on the neural correlates of subsequent memory. J Neurosci 24:4912–4917. CrossRef Medline

Mitchell JP, Cloutier J, Banaji MR, Macrae CN (2006) Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. Soc Cogn Affect Neurosci 1:49–55. CrossRef Medline

Noonan MP, Mars RB, Rushworth MF (2011) Distinct roles of three frontal cortical areas in reward-guided behavior. J Neurosci 31:14399–14412. CrossRef Medline

O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr Opin Neurobiol 14:769–776. CrossRef Medline

O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35–53. CrossRef Medline

Reeder GD (1993) Trait-behavior relations and dispositional inference. Pers Soc Psychol Bull 19:586–593. CrossRef

Reeder GD (2006) From trait-behavior relations to perceived motives: an evolving view of positivity and negativity effects in person perception. Polish Psychol Bull 37:191–202.

Reeder GD, Brewer MB (1979) A schematic model of dispositional attribution in interpersonal perception. Psychol Rev 86:61–79. CrossRef

Reeder GD, Coovert MD (1986) Revising an impression of morality. Soc Cogn 4:1–17. CrossRef

Reeder GD, Fulks JL (1980) When actions speak louder than words: implicational schemata and the attribution of ability. J Exp Soc Psychol 16:33–46. CrossRef

Reeder GD, Spores JM (1983) The attribution of morality. J Pers Soc Psychol 44:736–745. CrossRef

Reeder GD, Messick DM, Van Avermaet E (1977) Dimensional asymmetry in attributional inference. J Exp Soc Psychol 13:46–57. CrossRef

Rolls ET (2000) The orbitofrontal cortex and reward. Cereb Cortex 10:284–294. CrossRef Medline

Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and reward-guided learning and decision-making. Neuron 70:1054–1069. CrossRef Medline

Satpute AB, Badre D, Ochsner KN (2013) Distinct regions of prefrontal cortex are associated with the controlled retrieval and selection of social information. Cereb Cortex. Advance online publication. Retrieved November 5, 2013. doi: 10.1093/cercor/bhs408. CrossRef Medline

Schiller D, Freeman JB, Mitchell JP, Uleman JS, Phelps EA (2009) A neural mechanism of first impressions. Nat Neurosci 12:508–514. CrossRef Medline

Skowronski JJ, Carlston DE (1987) Social judgment and social memory: the role of cue diagnosticity in negativity, positivity, and extremity biases. J Pers Soc Psychol 52:689–699. CrossRef

Skowronski JJ, Carlston D (1989) Negativity and extremity biases in impression formation: a review of explanations. Psychol Bull 105:131–142. CrossRef

Souza MJ, Donohue SE, Bunge SA (2009) Controlled retrieval and selection of action-relevant knowledge mediated by partially overlapping regions in left ventrolateral prefrontal cortex. Neuroimage 46:299–307. CrossRef Medline

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Tsukiura T, Cabeza R (2011) Shared brain activity for aesthetic and moral judgments: implications for the beauty-is-good stereotype. Soc Cogn Affect Neurosci 6:138–148. CrossRef Medline

Van Overwalle F, Labiouse C (2004) A recurrent connectionist model of person impression formation. Pers Soc Psychol Rev 8:28–61. CrossRef Medline

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. Neuroimage 18:293–309. CrossRef Medline

Wojciszke B (2005) Morality and competence in person and self-perception. Eur Rev Soc Psychol 16:155–188. CrossRef

Wojciszke B, Brycz H, Borkenau P (1993) Effects of information content and evaluative extremity on positivity and negativity biases. J Pers Soc Psychol 64:327–335. CrossRef

Wojciszke B, Bazinska R, Jaworski M (1998) On the dominance of moral categories in impression formation. Pers Soc Psychol Bull 24:1245–1257.

Young L, Saxe R (2009) An fMRI investigation of spontaneous mental state inference for moral judgment. J Cogn Neurosci 21:1396–1405. CrossRef Medline

Zahn R, Moll J, Paiva M, Garrido G, Krueger F, Huey ED, Grafman J (2009) The neural basis of human social values: evidence from functional MRI. Cereb Cortex 19:276–283. Medline