Social Cognition, Vol. 33, No. 3, 2015, pp. 211–226

NEURAL RESPONSES TO APPEARANCE-BEHAVIOR CONGRUITY

Brittany S. Cassidy
Brandeis University and Indiana University

Angela H. Gutchess
Brandeis University

Research evidences stronger reactions toward those whose behaviors seem consistent with appearance. To better understand the processes underlying appearance-behavior congruity effects, we assessed regions responding as a function of the congruity between visual (appearance) and abstract (behavior) cues. Using fMRI, trustworthy- and untrustworthy-looking faces were paired with positive, negative, or neutral behaviors. Approach judgments were stronger for congruent over incongruent targets, replicating prior work. Incongruent targets (e.g., untrustworthy face/positive behavior) elicited medial prefrontal (mPFC) and dorsolateral prefrontal (dlPFC) cortex activity more than congruent (e.g., untrustworthy face/negative behavior), suggesting processing incongruent targets requires additional mentalizing and controlled processing. Individual differences in enjoying interpersonal interactions negatively correlated with mPFC activity toward incongruent over congruent targets, suggesting more effortful processing of incongruent targets for individuals with lower levels of social motivation. These findings indicate mPFC contributions to processing incongruent appearance-behavior cues, but suggest that individual differences may temper the extent of this effect.

This research was supported in part by an American Psychological Association Dissertation Research Award and a National Science Foundation Graduate Research Fellowship to B.S.C, and by Brandeis University. We thank Jonathan Jackson, Peter Millar, Laura Paige, and Tammy Hsieh for research assistance, and thank Kathryn Boucher and Anne Krendl for additional advice.

Address correspondence to Brittany S. Cassidy, Indiana University, Department of Psychological and Brain Sciences, 1101 E. 10th Street, Bloomington, IN 47405. E-mail: bscassid@indiana.edu.

Much behavioral (Uleman, Saribay, & Gonzalez, 2008; Zebrowitz & Montepare, 2008) and neuroimaging (Ma, Vandekerckhove, VanOverwalle, Seurinck, & Fias, 2011; Oosterhof & Todorov, 2008) work has explored the mechanics of impression formation. People associate actors with traits inferred from behaviors (Todorov & Uleman, 2002, 2003, 2004), and activation related to mentalizing (i.e., the process by which we make sense of others' mental states) underscores these associations (Ma et al., 2011). Behavioral research often controls for facial contributions to impressions (e.g., trustworthiness; Willis & Todorov, 2006), preventing an examination of how appearances impact behavior-based impressions. Overlooking facial information may simplify how we view the processes underlying impression formation given that people behave differently based on the congruity, or consistency, of social cues (for a review, see Stangor & McMillan, 1992). Identifying brain regions sensitive to appearance-behavior congruity can inform our understanding of the processes underlying impression formation and how appearance-behavior congruity leads to behavioral differences in thinking about others.

Defined by outcome dependency hypothesis (Erber & Fiske, 1984), people react more strongly to incongruent information when perceiver outcomes depend on the actions of an individual, but react more strongly to congruent cues when outcomes are not actor-dependent. Manipulating facial trustworthiness (which conveys valence; see Oosterhof & Todorov, 2008) and behavioral valence can vary appearance-behavior congruity. Research demonstrates advantages of congruent (e.g., trustworthy face/positive behavior) over incongruent (e.g., trustworthy face/negative behavior) cues when outcomes are not actor-dependent. For instance, source memory for undesirable headlines is best when provided by an untrustworthy- versus trustworthy-looking source (Nash, Bryer, & Schlaghecken, 2010). People remember more impressions of others (Cassidy, Zebrowitz, & Gutchess, 2012) and are more likely to approach people (Cassidy & Gutchess, 2014) with appearance-congruent versus incongruent behaviors.

Mentalizing could be differentially engaged based on appearance-behavior congruity. Forming impressions from incongruent appearance-behavior cues could require more mentalizing than when cues are congruent because resolving inconsistency requires resources (Macrae, Bodenhausen, Milne, & Wheeler, 1996; Macrae, Bodenhausen, Schloerscheidt, & Milne, 1999). The conflict present with incongruent cues may therefore *weaken* judgments toward others. People constantly associate behavior-inferred traits with actors, in activities ranging from perusing dating websites to reading the newspaper. Addressing neural responses to appearance-behavior congruity can complement behavioral work by clarifying the processes underlying decisions that vary by the consistency of social cues.

The present study advances the literature by assessing the neural response to appearance-behavior congruity established by abstract (behavior) *and* visual (appearance) cues. Medial prefrontal cortex (mPFC), a region widely involved in mentalizing (Ma et al., 2011), potentially underlies differences in processing congruent and incongruent appearance-behavior cues. For instance, mPFC engages when forming impressions of politicians paired with statements incongruent versus congruent with partisan beliefs (Cloutier, Gabrieli, O'Young, & Ambady, 2011). mPFC activity correlates with learning behaviors after first learning faces, suggesting a role in updating impressions over time (Baron, Gobbini, Engell, & Todorov, 2011). Impression updating work also revealed mPFC engagement, in addition to a general conflict network (Ma et al., 2012), when new behaviors contradict versus match behavioral expectations (Ma et al., 2012; Mende-Siedlecki, Cai, & Todorov, 2013). Multiple visual cues also evoke differential neural engagement, with stronger amygdala responses to incongruity based on expression and eye gaze (Adams, Gordon, Baird, Ambady, & Kleck, 2003), and mPFC engagement given expectancy violation from race and emotion cues (Hehman, Ingbretsen, & Freeman, 2014). We expected more mPFC activity for incongruent over congruent appearance-behavior cues, potentially reflecting more mentalizing to integrate and bind traits to actors.

Because social information influences visual processing (Adams, 2011; Eberhardt, Goff, Purdie, & Davies, 2004; Verosky, Todorov, & Turk-Browne, 2013), we predicted that unanticipated social cues could influence processing more than anticipated ones. Indeed, connectivity exists between prefrontal and ventral-visual regions when perceiving stereotype-incongruent versus congruent targets (Hehman et al., 2014; Quadflieg et al., 2011). We predicted activity within visual processing regions would functionally correlate with mPFC engagement for incongruent over congruent trials. However, given mPFC connectivity to dorsolateral prefrontal cortex (dlPFC) under trait inconsistency (Ma et al., 2012), we alternatively predicted connectivity between mPFC and dlPFC, a region critical in regulation and impression updating (Mende-Siedlecki, Cai, & Todorov, 2013).

Although studies have identified regions underlying differential processing of incongruent over congruent social cues, no work has considered *who* may most likely differ in their processing of incongruent and congruent information. Our secondary and more exploratory goal was to identify individual differences associated with processing incongruent over congruent social cues. Neuroimaging provides an ideal method, as individual differences in neural processing could result in similar behavior. For instance, young and older adults similarly evaluate social stigma, but activate brain regions differently to arrive at these evaluations, reflecting strategies compensating for age-related cognitive decline (Krendl, Heatherton, & Kensinger, 2009).

One characteristic potentially associated with processing social cues varying in congruity is extraversion, the extent to which people seek out and are comfortable interacting with new people. This social dimension of personality (Eysenck, 1967) has been linked to social competence (Schneider, Ackerman, & Kanfer, 1996) and decoding of nonverbal behaviors in contexts requiring cognitive control (Lieberman & Rosenthal, 2001). Examining differences on this characteristic may be promising in the cognitively demanding neuroimaging environment (Gutchess & Park, 2006). Individuals who engage in social interaction less might find processing incongruent over congruent targets more effortful, exhibiting increased mentalizing-related activity towards these targets. However, people who engage in interpersonal interaction more might also more likely spontaneously mentalize about others to learn about strangers.

METHOD

PARTICIPANTS

Eighteen right-handed adults with no history of neurological problems (18–34 years, 10 female; *M* = 26.24, *SD* = 5.36) from the Boston area participated and provided informed consent. Power analyses using an estimated medium effect size (*d* = .5) and 168 observations per participant indicated 16 participants to obtain power = .80 (Westfall, Kenny, & Judd, 2014). The Brandeis IRB approved this study.

STIMULI

Faces. Eighty-four trustworthy- and 84 untrustworthy-looking faces were selected from 740 publically available (http://tlab.princeton.edu/databases/) male faces generated by FaceGen version 3.1 (http://facegen.com). Faces were selected based on trustworthiness ratings (1 = very untrustworthy to 9 = very trustworthy) from 10 young (*M* age = 18.90 years, *SD* = 1.20) and 10 older (*M* age = 80.80 years, *SD* = 5.16) adults. Ratings were standardized by each age group's mean rating of all faces. Standardized ratings were entered into a 2 (Age Group: young, old) x 2 (Facial Trustworthiness: trustworthy, untrustworthy) ANOVA. Trustworthy faces (*M* = .79, *SD* = .48) were more trustworthy than untrustworthy (*M* = -1.06, *SD* = .65), $F(1, 332) = 876.62$, $p < .001$, $\eta_p^2 = .73$. Other effects were non-significant, *p*s > .70.

Behaviors. Fifty-six positive, 56 neutral, and 56 negative behaviors were selected from a dataset of 185 behaviors (Somerville, Wig, Whalen, & Kelley, 2006). Behaviors were selected based on arousal and valence ratings of eight older adults and eight of 23 younger raters (see Somerville et al., 2006). We averaged the young and older adults' ratings, and standardized ratings based on the mean rating of the dataset. Ratings were entered into ANOVAs comparing Category (positive, neutral, negative) on arousal, valence, and valence extremity (i.e., how strongly a behavior feels positive or negative). Tukey comparisons characterized differences.

The behaviors differed on arousal, $F(2, 165) = 312.92$, $p < .001$. Positive (*M* = .67, *SD* = .50) and negative (*M* = .65, *SD* = .48) were more arousing than

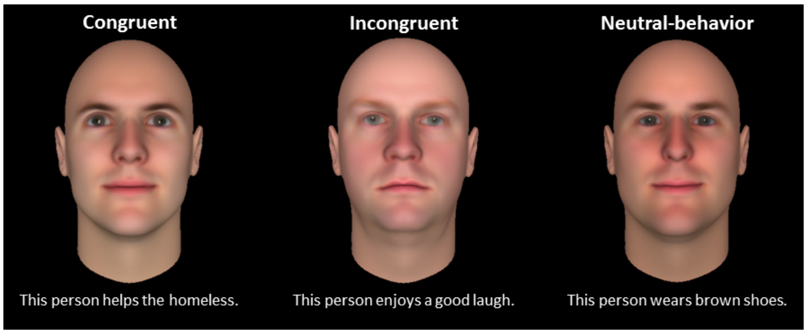

FIGURE 1. Examples of congruent, incongruent, and neutral-behavior targets.

neutral (*M* = -1.25, *SD* = .41), *p*s < .001. Positive and negative did not differ, *p* = .97. All of the categories differed from each other on valence, *F*(2, 165) = 1597.53, *p* < .001: positive (*M* = 1.18, *SD* = .25), neutral (*M* = .08, *SD* = .13), and negative (*M* = -1.21, *SD* = .27), *p*s < .001. The categories differed on valence extremity, *F*(2, 165) = 453.79, *p* < .001. Positive (*M* = .65, *SD* = .45) and negative (*M* = .70, *SD* = .59) had stronger ratings than neutral (*M* = -1.29, *SD* = .18), *p*s < .001. Positive and negative did not differ, *p* = .81.

Face-Behavior Pairs. Positive, negative, and neutral behaviors were equally distributed among trustworthy and untrustworthy faces. Trustworthy face/positive behavior and untrustworthy face/negative behavior pairs were categorized as "congruent." Trustworthy face/negative behavior and untrustworthy face/positive behavior pairs were categorized as "incongruent." Faces paired with neutral behaviors were categorized as "neutral-behavior pairs" (Figure 1). As the neutral-behavioral pairs were not directly relevant to our hypotheses, data involving these pairs will not be discussed further. Three versions, in which each face was paired with a positive, negative, or neutral behavior, counterbalanced the 168 pairs.

PROCEDURE

Stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Trials lasted 6000msec. Faces appeared alone for 2000msec and together with behaviors for 4000msec, during which participants were instructed to indicate their likelihood of approaching each person based on the given information (8-point scale; 1 = definitely would not; 8 = definitely would).

Data were collected over four seven-minute (210 TR) runs. Runs consisted of 42 pairs. Equal numbers of congruent, incongruent, and neutral-behavior pairs were viewed in each run. Trustworthy- and untrustworthy-looking faces were equally represented within all conditions in each run. Periods of baseline (indicated by a fixation cross) introduced jitter (Dale & Buckner, 1997) into the event-related design. Fixation intervals obtained using Optseq (http://surfer.nmr.mgh.harvard.edu/opt-seq) ranged from 0-22000msec.

Participants completed a post-task questionnaire including items related to task difficulty (e.g., "Did you find it more difficult to make decisions about people whose facial characteristics did not seem to match their behaviors?") and enjoyment of interpersonal interactions ("Do you feel comfortable approaching new people?"; "Do you like meeting new people?"; "How many close friends do you have?"; "How introverted or extroverted do you consider yourself to be?"; and "Do you like learning new things about strangers?"). All items used a 7-point scale (e.g., 1 = not at all to 7 = very much so).

IMAGE ACQUISITION AND ANALYSIS

Data were collected at the Center for Brain Science at Harvard University with a Siemens Trio 3.0T whole-body scanner using a 12-channel headcoil and an echo-planar imaging sequence (TR = 2000msec, TE = 25msec, FOV = 216mm, flip angle = 90) to acquire 34 slices 3.0mm thick with no skip and 3.0mm isotropic voxels. Anatomical images were acquired with a multi-echo MPRAGE sequence (176 slices, 1.0mm thick, TR = 2530msec, TE_1 = 1.64msec, TE_2 = 3.5msec, TE_3 = 5.36msec, TE_4 = 7.22msec) (Wonderlick et al., 2009).

Analyses were conducted in SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). Images were slice-time corrected, realigned to correct for motion, normalized to the MNI (Montreal Neurological Institute) template, and smoothed using an 8mm FWHM isotropic Gaussian kernel. We modeled events for the four-second epoch when faces and behaviors were displayed together. This epoch is an appropriate choice given that we wanted to model activity when appearance-based impressions were either confirmed or violated by behavior information. The two-second epoch when faces were viewed alone was modeled separately so these timepoints were not included in the implicit baseline. Because viewing faces alone preceded face-behavior pairings, it is inappropriate to consider neural responses to viewing faces alone. Thus, events (congruent, incongruent, or neutral-behavior pairs, and faces alone) were convolved with a canonical hemodynamic response function and used to create contrasts of interest.

General Linear Model. We assessed regions responsive to incongruent versus congruent pairs by contrasting [incongruent > congruent] over the four seconds in which faces and behaviors appeared together. First-level images were submitted to second-level one-sample t-tests, as in Cloutier and colleagues, 2011. Peak coordinates were identified by an extent threshold of 19 contiguous voxels exceeding a voxel-wise threshold of $p < .001$. One thousand Monte Carlo simulations indicated this provided a corrected experiment-wise threshold of $p < .05$. This method (described in Slotnick, Moo, Segal, & Hart, 2003) has been used in numerous fMRI studies of social cognition (Quadflieg et al., 2009). Brodmann areas were obtained with MRIcron (Rorden & Brett, 2000).

Psychophysiological Interaction. A psychophysiological interaction (PPI) analysis (Friston et al., 1997), performed using gPPI (http://www.nitrc.org/projects/gppi), examined functional connectivity to an mPFC seed identified at the group level GLM [incongruent > congruent] contrast (peak MNI coordinate: 6 50 1). gPPI accommodates more than two task conditions and simulations suggest greater sensitivity and specificity than SPM implementation (McLaren, Ries, Xu, & Johnson, 2012). The deconvolved time series from a 6mm radius sphere around this coordinate was extracted from each participant. First-level images from the PPI analysis were entered into a one-sample t-test to identify regions eliciting correlated activity with mPFC during incongruent versus congruent trials.

RESULTS

BEHAVIOR

Because our aim was to investigate neural responses based on appearance-behavior congruity, we verified a behavioral congruity effect also existed. We were interested in the relative judgment strength of congruent and incongruent pairs, as prior work identified stronger judgments toward congruent versus incongruent face-behavior pairs (Cassidy & Gutchess, 2014). Approach judgment strength, or the degree of approach motivation relative to average approach motivation, may be of interest in exploring congruity. Although people may approach trustworthy- and untrustworthy-looking people performing positive behaviors, congruity impacts judgment *strength*.

For each participant, the approach judgment for each pair was subtracted from his or her mean judgment. To compare across congruent (i.e., trustworthy-positive and untrustworthy-negative) and incongruent (i.e., trustworthy-negative and untrustworthy-positive) pairs, we took the absolute values of these subtractions. This accounts for individual differences in scale use and has been used in work considering social judgment and trait attribution strength (Cassidy & Gutchess, 2014; Follett & Hess, 2002). We collapsed across valence because behavioral work has converged on the idea that faces are better distinguished when paired with congruent versus incongruent behaviors (Cassidy & Gutchess, 2014; Rule, Slepian, & Ambady, 2012) and because we did not have a priori predictions regarding valence.[1] We used multi-level modeling (Raudenbush & Bryk, 2002) to

---

[1] Exploratory analyses showing congruity effects regardless of valence verified collapsing across valence was appropriate. For negative behaviors, stronger judgments toward untrustworthy- versus trustworthy-looking faces emerged. For positive, stronger judgments toward trustworthy- versus untrustworthy-looking faces emerged. Collapsing across valence allows us to most parsimoniously address our hypotheses.

TABLE 1. Brain Regions Responsive to Appearance-Behavior Congruity

A. Incongruent > Congruent

BA	**Incongruent > Congruent**	***k***	***t***	***p-value***	***x***	***y***	***z***
9/46	L dorsolateral prefrontal cortex	30	5.54	< .001	-30	26	37
10	R medial prefrontal cortex	34	5.21	< .001	6	50	1

B. Congruent > Incongruent

BA	**Congruent > Incongruent**	***k***	***t***	***p-value***	***x***	***y***	***z***
17	R primary visual cortex	244	7.70	< .001	15	-88	-2
18	L visual association area	187	6.01	< .001	-15	-97	1
18	L visual association area		5.47	< .001	-33	-85	-11
19	L extrastriate cortex		4.61	< .001	-45	-79	-5
19	R extrastriate cortex	49	4.62	< .001	42	-76	-11

Note. Regions listed without cluster sizes are subsumed by the cluster listed above. Regions are listed from highest to lowest t-value. L = left; R = right; k = cluster size; BA = approximate Brodmann's area; x, y, and z represent MNI coordinates of the peak voxel within each cluster. Cluster sizes are reported using a voxel-wise threshold of $p < .001$ and a 19-voxel extent threshold for an experiment-wise threshold of $p < .05$.

examine data where trials were nested within participants. Congruent and neutral-behavior pairs were dummy coded, such that incongruent pairs were an implicit baseline. The equations were as follows:

$$\text{Level 1: JUDGMENTij} = \beta0j+\beta1j^{*}(\text{CONGRUENTij})+\beta2j^{*}(\text{NEUTRAL\_BEHAVIORij})+\text{rij}$$

$$\text{Level 2: } \beta0j = \gamma00+u0j;\ \beta1j = \gamma10;\ \beta2j = \gamma20$$

$$\text{Mixed model: JUDGMENTij} = \gamma00+\gamma10^{*}\text{CONGRUENTij}+\gamma20^{*}\text{NEUTRAL\_BEHAVIORij}+u0j+rij$$

An effect of congruent face-behaviors emerged ($\beta_1 = .54$, $t(2942) = 12.66$, $p < .001$). Congruent pairs ($M = 2.50$, $SD = .45$) received stronger judgments relative to the mean approach tendency than incongruent ($M = 1.96$, $SD = .49$). An effect of neutral-behavior pairs also emerged ($\beta_2 = -.77$, $SE = .04$, $t(2942) = -18.29$, $p < .001$), indicating weaker judgments of neutral-behavior pairs ($M = 1.19$, $SD = 0.38$) relative to the mean approach tendency than incongruent. These comparisons implicitly suggest stronger judgments for congruent over neutral-behavior pairs.

MANIPULATION CHECK

We hypothesized that forming impressions from incongruent cues requires more mentalizing than from congruent given the resource demands of inconsistency resolution (e.g., Macrae et al., 1999), with conflict weakening judgments. If this is true, difficulty in making decisions about incongruent versus congruent targets should lead to more disparity between congruent and incongruent targets judgment strength. We calculated disparity for

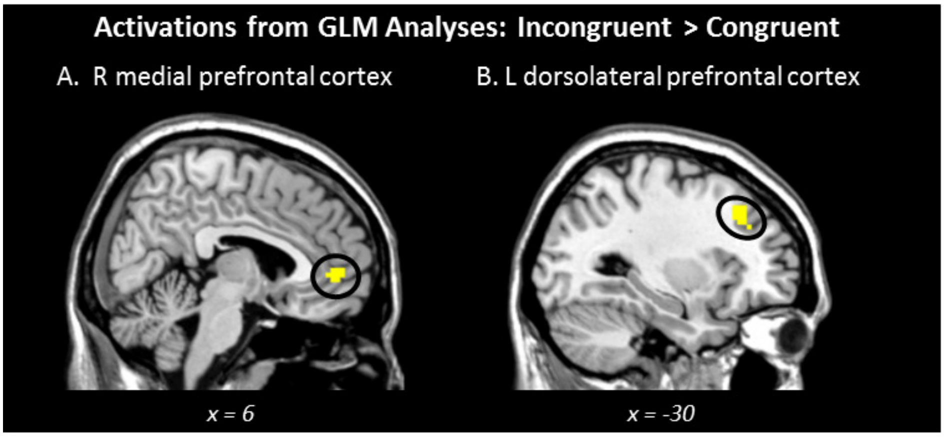

FIGURE 2. mPFC (A) and dlPFC (B) activated more for incongruent > congruent targets.

each participant by subtracting the mean judgment strength of incongruent targets from the mean of congruent. We then correlated disparity with self-reported difficulty of making decisions about incongruent over congruent targets. A correlation between disparity and difficulty emerged, $r = .70$, $p < .001$. Participants with more disparity between congruent and incongruent target judgment strength reported more difficulty in making decisions about incongruent targets.

IMAGING

GLM. We identified regions from the [incongruent > congruent] contrast to assess whether increased mentalizing supported the incongruity's effect on judgment strength. As hypothesized, greater recruitment of right mPFC existed (Table 1A, Figure 2A), consistent with related work (Cloutier et al., 2011; Hehman et al., 2014; Ma et al., 2012; Mende-Siedlecki, Cai, & Todorov, 2013) and supporting the idea that people mentalize more in response to incongruent social cues. Additionally consistent with prior work (Hehman et al., 2014; Ma et al., 2012), a region of left dlPFC emerged (Table 1A, Figure 2B). DlPFC engages in tasks requiring cognitive control (Miller & Cohen, 2001). Contrasting [congruent > incongruent] yielded widespread activity in visual processing regions (Table 1B), suggesting perhaps more attention to faces for anticipated behaviors.

Psychophysiological Interaction (PPI). PPI identified regions functionally connected with the mPFC seed for incongruent over congruent targets. No activations emerged at $p < .001$. Activity was coupled with dlPFC (BA 9; k = 9, $t = 4.26$, $p < .001$, peak MNI coordinates: 18 44 25) using a more liberal threshold. ($p < .005$; commonly utilized in related work, see Mende-Siedlecki, Baron, & Todorov, 2013). This suggests mPFC and dlPFC communicate more for incongruent versus congruent targets. mPFC-dlPFC

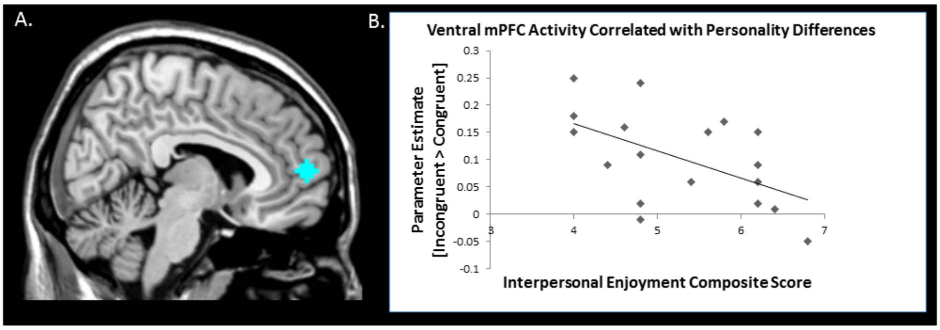

FIGURE 3. Increased activation to incongruent > congruent targets was evident in an independent ventral mPFC ROI (A). Composite interpersonal enjoyment scores correlated with ventral mPFC engagement to incongruent > congruent targets (B).

connectivity has been identified during social incongruity during spontaneous, but not intentional, impression formation (Ma et al., 2012). Our effect may be less robust than previous findings because our task involved approach ratings directly related to the intentional formation of impressions.

ROI and Individual Differences. We verified that mPFC response characterized activity related to the inconsistency of appearance-behavior cues using independent region of interest (ROI) analyses. We selected two mPFC regions (MNI coordinates: dorsal mPFC: 4 46 28, ventral mPFC: 6 56 10) identified in related work (Ma et al., 2012) as sensitive to violations in behavior-based trait expectations by opposite valence behaviors. Spherical ROIs of 6mm were defined based on MNI coordinates. At $p < .001$, activity in the [incongruent > congruent] contrast was identified in the ventral, but not dorsal, mPFC ROI (Figure 3A). Thus, our findings complement work on processing inconsistent trait-related information, extending this work to include processing appearance-behavior inconsistencies.

Our secondary goal considered individual differences contributing to processing incongruent over congruent appearance-behavior cues. Specifically, we were interested in how differences in introversion and interpersonal ease relate to processing inconsistent person information. Because five post-task questionnaire items (see Methods) related to seeking out interpersonal interactions and had high internal consistency (Cronbach's alpha = .83), we created a composite interpersonal enjoyment score by averaging these responses for each participant. We correlated [incongruent > congruent] parameter estimates from the ventral mPFC ROI with interpersonal enjoyment scores. A negative relationship emerged, $r = -.55$, $p = .02$ (Figure 3B). Those with lower interpersonal enjoyment had greater mPFC responses to incongruent over congruent targets. Those with lower interpersonal enjoyment did not, however, exhibit more disparity in judgment strength between congruent and incongruent targets, $r = .23$, $p = .35$.

DISCUSSION

Appearances elicit social expectancies (Ambady & Rosenthal, 1993; Olivola & Todorov, 2010). A core tenant of impression formation theory is that perceivers expect consistency in an individual's traits and behavior (Hamilton & Sherman, 1996). Expecting consistency suggests that processing information incongruent with appearances is cognitively effortful, an idea supported by behavioral work (Macrae, Bodenhausen, & Milne, 1994; Macrae et al., 1999), but crucial to maintain an organized impression. Forming impressions of people whose behaviors conflict (versus cohere) with appearance-based expectations may require increased mentalizing, with conflicting appearance-behavior cues ultimately tempering judgment strength. The present study used fMRI to inform behavioral work by clarifying the processes underlying stronger judgments toward congruent versus incongruent face-behavior pairs. Greater mPFC and dlPFC responses existed for incongruent over congruent targets, suggesting increased mentalizing and control when encountering incongruent appearance-behavior information.

MPFC activation is widely implicated in mentalizing (Ma et al., 2011; Van Overwalle, 2009). Thus, increased activity toward incongruent over congruent pairs could reflect more mentalizing to *integrate* social cues coming from separable domains, here visual (appearances) and abstract (behaviors). Sensitivity of mPFC to the congruency of appearance-behavior cues extends prior work finding differential neural activation based on inconsistent visual cues (Adams et al., 2003; Hehman et al., 2014), inconsistent abstract information (Cloutier et al., 2011), and for updating impressions based on inconsistent behaviors (Ma et al., 2012; Mende-Siedlecki, Cai, & Todorov, 2013). Our findings suggest a role for mPFC in integrating cues from multiple domains to form impressions, dovetailing nicely with recent work showing a causal role for dmPFC in impression integration (Ferrari et al., 2014). Moreover, mPFC activity complements behavioral models of person perception, in that when information about others violates our expectations of them, additional inferences and processing help resolve the violation (Brewer, 1988; Fiske & Neuberg, 1990).

Interestingly, applying stereotypes also elicits mPFC activity (Mitchell, Ames, Jenkins, & Banaji, 2009; Quadflieg et al., 2011; Quadflieg et al., 2009). A distinction between this work and ours is that our participants *integrated* behaviors with appearances, whereas stereotype work often involves deciding how much someone would act in a stereotyped way. Future work could investigate potentially differential mPFC contributions to integrating social information and stereotyping, and if activation reflects similar or different processes. For instance, multivariate pattern analysis could distinguish whether different mPFC subregions support these processes, which would indicate that these psychological processes are distinct from each other.

Encountering incongruent over congruent targets also elicited dlPFC activity. DlPFC supports cognitive control operations (Braver, Paxton, Locke, & Barch, 2009; for a review, see Miller & Cohen, 2001) potentially necessary to integrate incongruent appearance-behavior cues into organized impressions. Supporting this idea, related work has reported lateral prefrontal recruitment for expectation violating social stimuli (Cloutier et al., 2011; Ma et al., 2012; Weissman, Perkins, & Woldorff, 2008; Zaki, Hennigan, Weber, & Ochsner, 2010). More dlPFC activity when encountering incongruent over congruent targets mirrors classic social congruity work showing unexpected behaviors as more difficult to comprehend and requiring more effort to link together, increasing their prominence in observers' minds (Hastie & Kumar, 1979; Srull, 1981; Srull & Wyer, 1989).

It is therefore interesting that incongruity *tempered* judgment strength, when violations have exacerbated extremity in other work. Notably, work finding exacerbated extremity has involved categorical stereotypes (e.g., gender) and expectation-violating behaviors (Bettencourt, Dill, Greathouse, Charlton, & Mulholland, 1997). Facial trustworthiness is an individuating, versus categorical, distinction (Todorov, 2008). Reduced judgment strength given incongruent appearance-behavior cues could thereby better relate to work illustrating that congruent facial features strongly influence decision overconfidence (Hassin & Trope, 2000) as compared to work involving the violation of categorical stereotypes.

Our more exploratory goal regarded whether individual differences in interpersonal engagement influence the processing of incongruent versus congruent social cues. We provide initial evidence that less interpersonal engagement is linked to greater mPFC activity toward incongruent over congruent targets. Our findings support work demonstrating that introverted individuals exhibit non-verbal decoding deficits under cognitive load (Lieberman & Rosenthal, 2001). Such work is relevant to the present task, as successful decoding of non-verbal behavior predicts many social outcomes (Ambady & Rosenthal, 1992). Speculatively, less engagement in and enjoyment of interpersonal interactions could make processing incongruent over congruent targets more effortful, eliciting greater mentalizing. Indeed, relative introversion corresponds to myriad neuroanatomical structural (Wright et al., 2006) and functional (Canli, 2004; Gray et al., 2005) variations. Our measure of judgment strength disparity did *not* correlate with interpersonal enjoyment, consistent with work showing that differential neural processing can lead to similar behavior (Krendl et al., 2009) and highlighting a unique role for fMRI in elucidating individual differences in behavior. Albeit intriguing, this finding is only a first step in linking personality differences to mentalizing-related activity. Future research should utilize well-validated standard personality measures to clarify a personality mechanism underlying differential neural response to social congruity.

These findings inform past behavioral work (Cassidy & Gutchess, 2014; Cassidy et al., 2012; Kleider, Cavrak, & Knuycky, 2012; Nash et al., 2010;

Rule et al., 2012; Rule, Tskhay, Freeman, & Ambady, 2014; Suzuki & Suga, 2010) by localizing brain regions whose engagement underlies the processing of incongruent versus congruent appearance-behavior information. MPFC and dlPFC activity co-occur with congruity effects in judgment strength, connecting brain function linked to mentalizing and control to a key social behavior dependent on appearance-behavior congruity: approach likelihood. Moreover, individuals with *less* interpersonal ease activate mPFC *more* when encountering incongruent over congruent targets, suggesting differences in how people feel about social interaction impact brain activity supporting the processing of different social cues. Although future work can clarify specific individual differences and mechanisms supporting the processing of social cue congruity, these findings extend previous work (Cloutier et al., 2011; Ma et al., 2012; Mende-Siedlecki, Cai, & Todorov, 2013; Zaki et al., 2010) by demonstrating sensitivity to appearance-behavior congruity.

REFERENCES

Adams, R. (Ed.). (2011). *The science of social vision* (Vol. 7). New York: Oxford University Press.

Adams, R., Gordon, H., Baird, A., Ambady, N., & Kleck, R. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*, 1536.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256-274.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*(3), 431-441.

Baron, S., Gobbini, M., Engell, A., & Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social, Cognitive, & Affective Neuroscience, 6*, 572-581.

Bettencourt, B., Dill, K., Greathouse, S., Charlton, K., & Mulholland, A. (1997). Evaluations of ingroup and outgroup members: The role of category-based expectancy violation. *Journal of Experimental Social Psychology, 33*(3), 244-275.

Braver, T., Paxton, J., Locke, H., & Barch, D. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences, 106*(18), 2009.

Brewer, M. (1988). *A dual process model of impression formation*. Hillsdale, NJ: Erlbaum.

Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: Learning from individual differences in emotion processing. *Journal of Personality, 72*(6), 1105-1132.

Cassidy, B., & Gutchess, A. (2014). Influences of appearance-behavior congruity on memory and social judgments. *Memory*. Advance online publication. doi:10.1080/09658211.2014.951364

Cassidy, B., Zebrowitz, L., & Gutchess, A. (2012). Appearance-based inferences bias source memory. *Memory and Cognition, 40*(8), 1214-1224.

Cloutier, J., Gabrieli, J., O'Young, D., & Ambady, N. (2011). An fMRI study of violations of social expectations: When people are not who we expect them to be. *Neuroimage, 57*(2), 583-588.

Dale, A., & Buckner, R. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5*, 329-340.

Eberhardt, J., Goff, P., Purdie, V., & Davies, P. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*(6), 876-893.

Erber, R., & Fiske, S. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology, 47*(4), 709-726.

Eysenck, H. (1967). *The biological basis of personality*. Springfield, IL: Thomas.

Ferrari, C., Lega, C., Vernice, M., Tamietto, M., Mende-Siedlecki, P., Vecchi, T.,...Cattaneo. (2014). The dorsomedial prefrontal cortex plays a causal role in integrating social impressions from faces and verbal descriptions. *Cerebral Cortex*. [e-pub ahead of print.]

Fiske, S., & Neuberg, S. (1990). A continnuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1-74.

Follett, K., & Hess, T. (2002). Aging, cognitive complexity, and the fundamental attribution error. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57B*(4), 312-323.

Friston, K., Buchel, C., Fink, G., Morris, J., Rolls, E., & Dolan, R. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6*, 218-229.

Gray, J., Burgess, G., Schaefer, A., Yarkoni, T., Larsen, R., & Braver, T. (2005). Affective personality differences in neural efficiency confirmed using fMRI. *Cognitive, Affective, and Behavioral Neuroscience, 5*(2), 182-190.

Gutchess, A., & Park, D. (2006). fMRI environment can impair memory performance in young and elderly adults. *Brain Research, 1099*, 133-140.

Hamilton, D., & Sherman, S. (1996). Perceiving persons and groups. *Psychological Review, 103*(2), 336-355.

Hassin, R. R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*(5), 837-852.

Hastie, R., & Kumar, P. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*(1), 25-38.

Hehman, E., Ingbretsen, Z., & Freeman, J. (2014). The neural basis of stereotypic impact on mulitple social categorization. *Neuroimage, 101*, 704-711.

Kleider, H., Cavrak, S., & Knuycky, L. (2012). Looking like a criminal: Stereotypical black facial features promote face source memory error. *Memory and Cognition, 40*(8), 1200-1213.

Krendl, A., Heatherton, T. F., & Kensinger, E. (2009). Aging minds and twisting attitudes: An fMRI investigation of age differences in inhibiting prejudice. *Psychology and Aging, 24*(3), 530-541.

Lieberman, M., & Rosenthal, R. (2001). Why introverts can't always tell who likes them: Multitasking and nonverbal decoding. *Journal of Personality and Social Psychology, 80*(2), 294-310.

Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., & Fias, W. (2012). Inconsistencies in spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience, 7*(8), 937-950.

Ma, N., Vandekerckhove, M., VanOverwalle, F., Seurinck, R., & Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas. *Social Neuroscience, 6*(2), 123-138.

Macrae, C. N., Bodenhausen, G., & Milne, A. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*, 34-47.

Macrae, C. N., Bodenhausen, G., Milne, A., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition, 14*(1), 1-20.

Macrae, C. N., Bodenhausen, G., Schloerscheidt, A., & Milne, A. (1999). Tales of the unexpected: Executive function and person perception. *Journal of Personality and Social Psychology, 76*(2), 200-213.

McLaren, D., Ries, M., Xu, G., & Johnson, S. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage, 61*, 1277-1286.

Mende-Siedlecki, P., Baron, S., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *Journal of Neuroscience, 33*(50), 19406-19415.

Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience, 8*, 623-631.

Miller, E., & Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Mitchell, J., Ames, D., Jenkins, A., & Banaji, M. R. (2009). Neural correlates of stereotype application. *Journal of Cognitive Neuroscience, 21*(3), 594-604.

Nash, R., Bryer, O., & Schlaghecken, F. (2010). Look who's talking! Facial appearance can bias source monitoring. *Memory, 18*(4), 451-457.

Olivola, C., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*, 83-110.

Oosterhof, N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087-11092.

Quadflieg, S., Flannigan, N., Waiter, G., Rossion, B., Wig, G. S., Turk, D., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *Neuroimage, 57*(2), 549-557.

Quadflieg, S., Turk, D., Waiter, G., Mitchell, J., Jenkins, A., & Macrae, C. N. (2009). Exploring the neural correlates of social stereotyping. *Journal of Cognitive Neuroscience, 21*(8), 1560-1570.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*: Thousand Oaks, CA: Sage.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioral Neurology, 12*(4), 191-200.

Rule, N., Slepian, M., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125*, 207-218.

Rule, N., Tskhay, K., Freeman, J., & Ambady, N. (2014). On the interactive influence of facial appearance and explicit knowledge in social categorization. *European Journal of Social Psychology, 44,* 529-535.

Schneider, R., Ackerman, P., & Kanfer, R. (1996). To "act wisely in human relations": Exploring the dimensions of social competence. *Personality and Individual Differences, 15*, 469-481.

Slotnick, S., Moo, L., Segal, J., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research, 17*, 75-82.

Somerville, L., Wig, G., Whalen, P., & Kelley, W. M. (2006). Dissociable medial temporal lobe contributions to social memory. *Journal of Cognitive Neuroscience, 18*(8), 1253-1265.

Srull, T. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory, 7*(6), 440-463.

Srull, T., & Wyer, R. (1989). Person memory and judgment. *Psychological Review, 96*(1), 58-83.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111*(1), 42-61.

Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition, 117*, 224-229.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124*, 208-224.

Todorov, A., & Uleman, J. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83*(5), 1051-1065.

Todorov, A., & Uleman, J. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39*(6), 549-562.

Todorov, A., & Uleman, J. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87*(4), 482-493.

Uleman, J., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology, 59*, 329-360.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30*, 829-858.

Verosky, S., Todorov, A., & Turk-Browne, N. (2013). Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neurpsychologia, 51*, 2100-2108.

Weissman, D., Perkins, A., & Woldorff, M. (2008). Cognitive control in social situations: A role for the dorsolateral prefrontal cortex. *Neuroimage, 40*(2), 955-962.

Westfall, J., Kenny, D., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020-2045.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-Ms exposure to a face. *Psychological Science, 17*(7), 592-598.

Wonderlick, J., Ziegler, D., Hosseini-Varnamkhasti, P., Locascio, J., Bakkour, A., van der Kouwe, A.,...Dickerson, B. (2009). Reliability of MRI-derived cortical and subcortical morphometric measures: Effects of pulse sequence, voxel geometry, and parallel imaging. *Neuroimage, 44*, 1324-1333.

Wright, C., Williams, D., Feczko, E., Feldman-Barrett, L., Dickerson, B., Schwartz, C., & Wedig, M. (2006). Neuroanatomical correlates of extraversion and neuroticism. *Cerebral Cortex, 16*(12), 1809-1819.

Zaki, J., Hennigan, K., Weber, J., & Ochsner, K. (2010). Social cognitive conflict resolution: Contributions of domain-general and domain-specific neural systems. *Journal of Neuroscience, 30*(25), 8481-8488.

Zebrowitz, L., & Montepare, J. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2/3*, 1497-1517.