UNIVERSITY OF CALIFORNIA

Santa Barbara

Neural and Facial Correlates of Affective Disposition during Morally-Salient Narratives

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Communication

by

James Michael Mangus

Committee in charge:

Professor René Weber, chair

Professor Daniel Linz

Professor Scott Grafton

March 2016

The dissertation of James Michael Mangus is approved.

Daniel Linz

Scott Grafton

René Weber, Committee Chair

December 2015

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the tremendous support of my committee. I am particularly appreciative of René's boundless enthusiasm and inexhaustible patience over the many years we have now worked together. Dan has provided me an exemplary role-model as an academic who is open-minded and refuses to sacrifice his own interests or personality for the sake of conformity. Scott taught me most of what I know about doing fMRI analysis in FSL, as well as a great number of things I have surely forgotten, with excellent clarity and concision.

I also have to acknowledge the support of my family, who have not only tolerated that I live thousands of miles away and never call, but actively encouraged me to pursue my interests. I am especially indebted to my late grandfather, Nick Mangus, who passed while I was conducting this research. He understood that the best gift is always an interesting book, and provided me with scores of them over the years.

Finally, I am thankful to my cat, Felix, who was always around to comfort me when things went poorly and who provided a constant supply of adorable high-fives when they went well.

VITA OF JAMES MICHAEL MANGUS
December 2015

EDUCATION

Bachelor of Science in Information Science, University of Pittsburgh, 2010 (magna cum laude)
Master of Arts in Communication, University of California, Santa Barbara, 2013
Doctor of Philosophy in Communication, University of California, Santa Barbara, 2016 (expected)

PUBLICATIONS

Weber, R., Eden, A., Huskey, R., Mangus, J. M., & Falk, E. (2015). Bridging media psychology and cognitive neuroscience: Challenges and opportunities. *Journal of Media Psychology, 27*(3), 146-156.

Mangus, J. M., Adams, A., and Weber, R. (2015). Media neuroscience. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences.* Hoboken, NJ: Wiley.

Weber, R., Mangus, J. M., & Huskey, R. (2015). Brain imaging in communication research: A practical guide to understanding and evaluating fMRI studies. *Communication Methods and Measures, 9*(1-2), 5—29.

Weber, R., Huskey, R., Mangus, J. M., Westcott-Baker, A., & Turner, B. O. (2015). Neural predictors of message effectiveness during counterarguing in antidrug campaigns. *Communication Monographs, 82*(1), 4—30.

Pure, R. A., Markov, A. R., Mangus, J. M., Metzger, M. J., Flanagin, A. J., & Hartsell, E. (2013). Understanding and evaluating source expertise in an evolving media environment. In T. Takseva (Ed.), *Social software and the evolution of user expertise: Future trends in knowledge creation and dissemination.* Hershey, PA: IGI Global.

Weber, R., Popova, L., & Mangus, J. M. (2012). Universal morality, mediated narratives, and neural synchrony. In R.Tamborini (Ed.), *Media and the Moral Mind*, London: Routledge.

AWARDS

University of California Regents' Special Fellowship (2010-2015)

ABSTRACT

Neural and Facial Correlates of Affective Disposition during Morally-Salient Narratives

by

James Michael Mangus

The recent growth of the neurophysiological paradigm has re-defined important concepts in communication. To better understand of how fictional narratives operate at a neurophysiological level, this study employs a combination of fMRI and face-tracking to explore affective disposition theory (ADT), which predicts that viewers' affective responses lead them to prefer narratives in which virtuous characters are rewarded and immoral characters are punished. Previous work has shown that inter-subject correlations (ISC) in brain activity are highest when viewing disposition-consistent narrative outcomes -- specifically, the punishment of immoral characters. The present study partially replicates these findings and also uses psychophysiological interaction (PPI) analysis to augment the notion that punishment of immoral characters yields discernibly different patterns of brain connectivity than other narrative content. To directly address the affective component of ADT, further PPI analyses compared high- and low-empathy participants. Results indicate that the patterns of co-activation between brain regions revealed through PPI are moderated by trait empathy: seeing good characters rewarded yields the same co-activation patterns among high-empathy individuals as seeing bad characters punished does among low-empathy individuals.

To provide another window into affective processing, automated face-tracking is used to evaluate whether greater ISCs in brain activity also yield greater similarities in the time-course of emotive facial expressions. Results indicate that correlations in facial expression vary systematically by experimental condition, but, contrary to the pattern of neuronal ISC, facial expressions exhibit the greatest correlation in disposition-*inconsistent* conditions. Furthermore, unlike neuronal ISC, correlations in facial expressions are significantly higher among high-empathy participants. In sum, these results support the view that disposition-consistent narrative content drives inter-subject correlation in brain activity, but that shared brain activity does not yield correlated displays of emotion; instead, emotive displays are moderated by empathy and may play a communicative role in expressing dissatisfaction with disposition-inconsistent narratives. The implications of these findings for further research into the affective component of ADT are discussed.

Neural and Facial Correlates of Affective Disposition during Morally-Salient Narratives

A broad framework for understanding communication as interbrain coupling has emerged in recent years, most notably in the work of Hasson and colleagues (Hasson et al., 2004; Hasson et al., 2008; Hasson et al., 2012). According to this view, communication aligns brain states across individuals. There is evidence consistent with this view for both non-verbal and verbal communication. For instance, Goldman and Sripada (2005) promote a simulation model for emotional processing of facial expressions, which holds that an observer attempts to replicate the mental state of another person based on their facial expression. This view is similar to the "unifying view" of Gallese, Keysers, & Rizzolatti (2004), which argues that overlapping brain structures are engaged by both first-person and third-person experience. Similarly, Stephens, Silbert, and Hasson (2010) find that communication is associated with synchronous alignment of brain activity between the sender and receiver. They contend that stronger temporal coupling of the brain during communication is evidence of greater comprehension.

Communication involves many brain systems because a mental representation of a certain state of affairs integrates diverse cognitive and affective components (see Weber, Sherry, & Mathiak, 2008). Narratives in particular present a fascinating case of multi-level communication. They convey both literal and figurative meaning, and oftentimes the means of expression matters as much as what is expressed. Narratives can encode characters, conflicts, moral lessons, and even abstract aesthetic performances that convey meta-level messages about narratives themselves. They are polysemic, amenable to multiple interpretations depending on individual and cultural context, but these interpretations are always necessarily related to human capacities for encoding and decoding certain types of information.

Narrative structure itself appears to be deeply rooted in the brain. Ross (2007) goes so far as to argue that narrative selfhood is the unique defining characteristic of being human. According to his view, the organization of self-related information takes a narrative form which draws on our capacity for language. Fictional narratives in particular serve important socially-oriented functions as well. Literary fiction has been shown to improve performance in theory of mind tasks, for instance (Kidd & Castano, 2013). Furthermore, fictional narratives allow individuals to consider counterfactual scenarios, serving as a safe virtual laboratory for considering how to react in dangerous or controversial situations. In fact, one theory suggests that counterfactuals are represented as "structured event complexes" in the medial prefrontal cortex (mPFC; Barbey, Krueger, & Grafman, 2009), supporting the idea that the capacity for a narrative-like organization of events underlies humans' capacity for hypothetical reasoning.

To develop the brain-coupling view, communication scholars must identify the neuronal systems that are coupled by specific communicative acts. For example, Hasson et al. (2008) manipulate videos to alter the narrative structure and measure cortical activity through fMRI. The results indicate that different films induce different levels of whole-brain inter-subject correlation (ISC): when viewing an episode of *Alfred Hitchcock Presents,* participants showed significant levels of ISC across >65% of cortex; when viewing an episode of *Curb Your Enthusiasim,* only 18%. For videos with limited narrative structure, ISC is even lower. They take ISC as a measure of collective engagement -- the extent to which viewers share a collective experience in response to a mediated narrative -- and suggest that varying degrees of collective engagement can be induced by different media genres and narrative structures.

Communication scientists can lend important theoretical insight to guide interbrain coupling research by developing well-founded predictions about the most relevant narrative features. But this research paradigm is relatively new, and it is not yet well-established how high-level narrative features -- the conceptual building blocks of key theories in communication science -- modulate ISC. For example, Weber, Eden, & Mathiak (2011) manipulated the valence (moral/immoral) and outcome (reward/punishment) of narratives and found preliminary evidence that ISC in relevant brain regions was highest in the immoral-punishment condition. The high ISCs produced by those narratives suggests that the punishment of immoral people evokes a shared response grounded in fundamental intuitions. To re-interpret the findings of Hasson et al. (2008), it seems likely that Alfred Hitchcock's morally-laden narratives yield greater collective engagement because their content evokes commonly- and deeply-held moral beliefs.

This study extends and complements prior work by using neuroimaging and automated face-tracking to explore the emotional groundwork of disposition theory. It will proceed in three steps. First, I attempt to replicate the preliminary findings of Weber et al. (2011) by conducting an exploratory whole-brain analysis, as well as specific region of interest (ROI) analyses. Second, I use psychophysiological interaction analysis to explore how viewers' preferences for specific narrative content may be driven by functional connectivity between theoretically-relevant ROIs. Finally, computer-automated face-tracking is used to test whether correlations in the emotiveness of participants' facial expressions track correlations in their brain activity.

Disposition, Emotion, and the Cognitive Neuroscience of Morality

Narratives are especially remarkable for their emotional evocativeness. Since classical antiquity, philosophers have recognized the importance of emotional evocativeness in narratives. One traditional explanation for the evocative power of a narrative is willful suspension of disbelief -- the notion that individuals voluntarily opt to be mentally transported to the fictional world of the narrative. However, Zillmann (2006, 2013) critiques this view. Under his account, the representation in media of an emotion-inducing stimulus can activate the same adaptive brain structures as the actual (i.e., nonfictional and nonmediated) experience of that same stimulus, particularly if the mediated representation is iconic (as in film) rather than symbolic (as in literature).

The instinctive reactions of other animals to lifelike mediated stimuli illustrate this point clearly. If you place a screen with video of moving insects in the visual field of a frog, the frog will try to snag the insects with its tongue as though they were actually present in the environment. It makes more sense, under this evolutionary account, to say that our *disbelief* of mediated stimuli is likely more willful than our *belief* of them, since mediated environments still activate instinctual responses. Understanding that the stimulus is mediated seems to be governed by executive-level processes in other brain structures, such that the autonomic emotional reaction is distinct from volitional regulation of behavioral response.

To elaborate the conceptual groundwork for his view, Zillmann (2006) argues that emotional responses to media should be understood along three dimensions: disposition, excitation, and experience. Disposition captures the hedonic valence of the emotion (e.g. positive or negative); excitation is the level of arousal induced by an emotion; the experience of an emotion emerges from cognitive elaboration. In his account, the extent to which we desire to see a character rewarded or punished depends on our disposition toward that character, which is based on a moral judgment of that character's acts. Disposition theory has classical roots in Aristotle, who advised that in a narrative “a good man must not be seen passing from happiness into misery” (quoted in Zillmann, 2006). Zillmann's approach largely accords with the Aristotelian formulation: he holds that preference for a certain outcome is driven by emotion because viewers empathize with morally virtuous characters and expect that they will be rewarded (i.e. that empathizing with the character will produce positive hedonic valence).

Emotions seem to provide the basis for the complex moral rules that govern human culture. Emotions are “specific and consistent collections of physiological responses triggered by certain brain systems when the organism represents certain objects or situations” (Damasio, 2000, p. 15). In popular use, emotion and reason are seen as opposing forces – rational thought is contrasted with emotional intuition. However, the somatic marker hypothesis advanced by Damasio (2000) maintains that both conscious reasoning and preconscious emotional processes play an important role in decision-making. In Damasio’s terms, an emotion is a basic intuition, whereas a feeling is a more elaborated evaluation of that intuition. In line with that distinction, it appears that basic emotions interface with deliberate reasoning to produce complex moral evaluations. For example, the limbic system facilitates basic emotions like anger and disgust, and considering moral violations yields increased connectivity between the limbic system and the prefrontal cortex, orbitofrontal cortex, precuneus and superior temporal sulcus (Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005). Other work has found that both socially-oriented moral disgust and pathogen disgust produce limbic system activation, but sociomoral disgust yields significantly stronger activation in medial prefrontal cortex and near the temporal-parietal junction than pathogen-related disgust (Schaich-Borg, Lieberman, & Kiehl, 2008).

Disentangling the multiplicity of processing systems involved in moral reasoning is an ongoing area of research in moral psychology (Dinh & Lord, 2013). For instance, emotionally-laden deontic moral intuitions frequently override a more reasoned utilitarian cost-benefit analysis, and under circumstances where cost-benefit analysis is complicated by unknown outcomes, the strict application of deontic rules may in fact be superior to a purely consequentialist analysis (Bennis, Medin, & Bartels, 2010). In order to evoke these moral conflicts in participants, studies frequently present hypothetical moral scenarios based on the famous “trolley problem” (Thomson, 1976). This problem takes many forms, but generally brings deontic intuitions in conflict with consequentialism. For instance:

> Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing the five. (Thomson, 1976).

Different formulations of the trolley problem and differences in its contextual presentation often lead to divergent responses. Mikhail (2007) argues that different responses to the trolley problem are the product of different mental representations of the problem. Drawing on neuroimaging studies, Greene and colleagues have attributed those different mental representations primarily to differences in emotional processing (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Moral scenarios evoke emotional responses that are subject to selective cognitive elaboration to reach a final judgment, and deontic judgments indicate that emotional intuitions have overwhelmed consequentialist reasoning.

Most recently, Shenhav & Greene (2014) have found that integrative moral judgment depends specifically on connectivity between the amygdala, which drives the affective response, and ventromedial prefrontal cortex (vmPFC), which supports executive processing and emotion inhibition. Specifically, they conclude that amygdala-vmPFC connectivity is highest when participants are asked to evaluate their emotional response to a moral scenario, and connectivity is lowest when participants are asked to assess a scenario in purely calculative, utilitarian terms. The two regions work in concert to yield an all-things-considered moral judgment.

Neurocinematics of Moral Psychology

From a communication science perspective, the incorporation of neuroimaging provides an observational measure of the emotional component of moral processing to further develop disposition theory. As summarized above, Zillmann (2006) explains disposition theory as an extension of emotional processing. Fictional narratives frequently depict dangerous and morally-controversial situations which evoke automatic emotional responses. Viewers' dispositions toward characters influence preferences for narratives – people prefer narratives in which liked characters are rewarded and disliked characters are punished (Raney & Bryant, 2002; Weber, Tamborini, Lee, & Stipp, 2008; Zillmann and Cantor, 1977). Based on initial evidence from Weber et al. (2011), it seems that this preference is reflected in the brain: ISC should be highest in disposition-consistent scenarios, and lowest in disposition-inconsistent scenarios. The negatively-valenced disposition-consistent scenes – wherein immoral characters are punished for their wrongdoing – are of particular interest given the preference for altruistic punishment demonstrated in prior research: individuals readily punish others, including third-parties, for their misdeeds (Buckholtz et al., 2008), even if doing so comes at a cost to the punisher (Fehr & Gachter, 2002).

> H1: Inter-subject correlation (ISC) in ROIs associated with moral reasoning will be higher when viewing disposition-consistent scenes than disposition-inconsistent scenes, with the greatest ISC when immoral characters are punished.

It is also worthwhile to consider a more narrowly-tailored test. If Shenhav and Greene are correct that the amygdala and vmPFC work together to produce integrative moral judgment, then their shared activation across subjects should be linked to moral content in the narrative. Additionally, while Weber et al. considered several ROIs associated with various aspects of moral psychology, amygdala and vmPFC were not among them. The current study therefore attempts to replicate Weber et al.'s findings using these specific ROIs.

H2: Inter-subject correlation (ISC) in (a) amygdala and (b) vmPFC will be higher when viewing in disposition-consistent scenes than disposition-inconsistent scenes, with the greatest ISC when immoral characters are punished.

Although there are reasons to be bullish about the future of fMRI studies in this vein, it is only fair to consider alternative measures of emotionality. Even if the brain is ultimately the seat of cognition, brain imaging is not always the most efficient way to gather information about individuals' cognitive states. Consequently, this study considers an alternative to brain activity which can convey emotional disposition: facial expressions. A substantial body of research supports the view that facial expressions are crucial means of emotional communication to which humans are innately sensitive (see Adolphs, 2002 for a review of neurophysiological evidence). For more than a decade, researchers have been refining computational techniques to measure these features as indicators of emotion (e.g., Busso et al., 2004). This study uses automated content-analysis of facial expressions during think-aloud sessions to triangulate the emotional state of participants.

Given that facial expressions convey emotion, and that fictional narratives stimulate and entertain by virtue of their emotional evocativeness, the most intuitive prediction would be that the disposition-consistent narratives, which yield both greater enjoyment and greater neuronal ISC, will also yield the greatest inter-subject similarities in the time-course of facial expressions. Thus, mirroring the previous hypotheses:

H3: Correlations of the facial expressions between individuals will be higher when viewing disposition-consistent scenes than disposition-inconsistent scenes, with the greatest correlations when immoral characters are punished.

Predicting Individual Preferences

One shortcoming in the cognitive neuroscience of morality is the use of contrived, forced-choice moral scenarios. In a recent review, Avramova and Inbar (2013) observe strong evidence that emotions sway moral judgment, but suggest caution regarding the automaticity of moral judgment; it has not yet been firmly established to what extent emotional intuitions "moralize" the world or, conversely, to what extent moral reasoning is a motivated process (Haidt, 2003; Pizarro & Bloom, 2003). Studying the automaticity of moral judgment is complicated by operational and theoretical problems (Avramova & Inbar, 2013; Moll et al., 2005; Narvaez, 2010), but these problems can be somewhat ameliorated by using more naturalistic stimuli. While Shenhav and Greene are interested in using forced moral evaluations to map the resultant patterns of brain activity, the study proposed here also concerns the extent to which variations in those functional connectivity patterns predict viewers' preferences for naturalistic fictional narratives. The brain-as-predictor approach has shown great promise in studies of persuasion using naturalistic stimuli (Berkman & Falk, 2013), and it may be equally useful in neurocinematics (Hasson, 2008).

Dispositions depend on moral judgments regarding characters' actions, and those judgments will vary across individuals (Tamborini, 2011). If disposition theory is correct that emotional inputs drive the preference for disposition-consistent content, then between-subjects differences in amygdala-vmPFC connectivity – reflecting the integration of both emotional input and executive judgment – may correspond with differences in preferences for disposition-consistent outcomes. Specifically, higher connectivity should be associated with a stronger preference for disposition-consistent outcomes, whereas lower connectivity should indicate more deliberative processing that attenuates the emotional effects of narrative content.

In other words, functional connectivity between the amygdala and vmPFC may accentuate both how much individuals like disposition-consistent narrative outcomes and how much they dislike disposition-inconsistent narrative outcomes.

> H4: Individuals with greater functional connectivity between the amygdala and vmPFC will have a stronger preference for disposition-consistent outcomes than individuals with less amygdala-vmPFC connectivity.

If this relationship is borne out, it can serve as stepping stone for a broader re-analysis of the experiential aspect of Zillmann's theoretical framework, using a brain-as-predictor approach to model dramaturgical preferences. Although, in the aggregate, people generally prefer narrative content where virtuous characters are rewarded and evil characters are punished, there are many notable exceptions to this principle. Various narrative features or individual differences could yield lower amygdala-vmPFC connectivity and, in turn, attenuate the impact of emotional disposition on the viewer's overall experience of the narrative.

Finally, as an exploratory research question, additional analyses will consider whether trait empathy underlies individual differences in moral processing. Research frequently links moral judgment with empathic concern (e.g. Decety, Michalska, & Kinzler, 2012; Moll & de Oliveira-Souza, 2007; Singer et al., 2006). Previous research has shown that lower trait empathy modulates moral judgment – for instance, it attenuates moral engagement (Detert, Trevino, & Sweitzer, 2008) and is associated with more utilitarian reasoning (Gleichgerrcht & Young, 2013). Participants' self-reported trait empathy is therefore considered as a potential moderating variable.

Method

The data for this study were collected by the Media Neuroscience Lab and comprise fMRI imaging and think-aloud sessions to gauge the participants' reactions to the narratives using multiple measures. Participants (original n = 28, excluding subjects with unusable or missing data, final n = 22; 100% female) watched professionally edited, 180s clips from the soap opera *Days of Our Lives*. Clips were manipulated to vary in their moral content - both valence (moral/immoral) and outcome (reward/punishment) - as well as a fifth neutral, amoral condition. Each participant watched the same set of scenes in a randomized sequence during fMRI scanning, with a 30-second resting baseline period between scenes. BOLD contrast was obtained with a gradient-echo echo-planar imaging (EPI) sequence (General Electric scanner; field strength of 3 Tesla; whole brain coverage with 30 interleaved slices; slice size 4mm with 0.4mm gap; TR = 2000ms; TE = 27.2 ms; flip angle = 77°, field of view 22 × 22 cm2, matrix size 64 × 64). The same participants watched the same scenes outside the scanner in think-aloud sessions, which were videotaped with the participant's face and the stimulus video visible. Participants also rated their perceptions of characters' morality, outcome, and importance for each stimulus video, which were used to calculate an individual ADT index by summing the squared differences between morality and outcome valence for all characters, weighted by character importance (see Weber et al., 2008, which validates this index by predicting television ratings). Participants also completed a self-report battery to measure their level of trait empathy (Davis, 1983).

Analytical Procedure and Results

fMRI Analysis

Preprocessing. Raw DICOM data for each functional run were assembled into NIFTI-format files and preprocessed using the tools provided by FSL 5.0 (http://fsl.fmrib.ox.ac.uk/). Slice-timing correction and brain-extraction were applied to the functional data, as well as high-pass temporal filtering (210s cutoff) and spatial smoothing (FWHM 5mm). Motion correction was performed with MCFLIRT, and excess-motion parameters were saved for each subject to be used as confound EVs in subsequent analyses. Registration was carried out in two steps: first, linear registration was used to fit the functional data to the subject's own high-resolution anatomical scan (T1-weighted SPGR) using the boundary-based registration (BBR) algorithm (Greve & Fischl, 2009), and then non-linear registration was performed to register each subject's brain to the Montreal Neurological Institute 152 (MNI-152) template (2mm resolution).

ROI masks. Masks for ROI analysis were created using FSL's command line tools. Masks for left and right amygdala were anatomically defined using the Harvard-Oxford structural atlas. Additionally, 5mm spherical masks were created around the 4 sets of vmPFC coordinates reported by Shenhav and Greene (2014). These 4 regions consist of a pair of lateralized masks for each of their two analytical contrasts: integrative > utilitarian moral judgment, and integrative > emotional moral judgment. These two pairs of masks therefore represent different estimates of the vmPFC region activated by an all-things-considered integrative moral judgment, evoked when participants consider which action they "would find more morally acceptable" in a forced-choice moral scenario. The former pair of masks was generated from the peak coordinates identified when contrasting integrative judgment with a utilitarian prompt: which action "would produce better results." The latter pair was generated from the peak coordinates identified by contrasting integrative judgment with an emotional prompt: which action the participant would "feel worse about doing." Since the participants in the current study are engaged in naturalistic viewing that does not explicitly elicit either utilitarian or emotional judgment, and because it seems unlikely that these regions would be clearly functionally differentiated anyway, both pair of masks are used as seed regions for this analysis.

Finally, for exploratory analyses, masks were downloaded for each region in the theory of mind (ToM) network identified by MIT's Saxelab (Dufour et al., 2013; masks available at http://saxelab.mit.edu). The ToM network overlaps extensively with networks identified in prior studies of moral judgment, as is to be expected given the theoretical overlap between the two: understanding the moral content of a narrative requires the viewer to consider the beliefs and intentions of characters (Bzdok et al., 2012). The ToM regions of particular interest given their consistent association with moral judgment are the superior temporal sulcus (STS), temporal-parietal junction (TPJ), precuneus (PC), and medial PFC; specific masks are provided for left and right TPJ, as well as for right STS, dmPFC, mmPFC, and vmPFC.

For each mask, a functional time-series was extracted from each subject's preprocessed data using the fslmeans command. The values extracted from anatomically-defined masks were weighted based on the atlas-given probability map for the ROI;

functionally-defined masks from prior studies were treated as binary and thus not weighted. The resultant time-series were then used for the connectivity analyses described below.

ISC analysis. Preprocessed functional data for each subject was spliced by condition using FSL, and these inputs were used for a whole-brain ISC analysis with contrasts between conditions. Inter-subject correlation in cortical activity was calculated using the Matlab analysis toolbox provided by Kauppi, Pajula, and Tohka (2014), which implements a novel ISC significance-testing procedure based on the Pearson-Filon statistic.

This analysis revealed numerous regions with significant inter-subject correlation across all conditions (see Figure 1). These regions include some obvious structures - for instance, the significant ISC in occipital cortex near the calcarine sulcus (max r = 0.15, FDR-corrected p<0.001, MNI-152 coordinates 8,-86,0) is no surprise because subjects all watched the same stimulus videos and this region is well-known for its role in processing visual stimuli. The same reasoning explains the area of highly-significant ISC in the vicinity of auditory cortex (max r = 0.17, FDR-corrected p < 0.001, MNI-152 coordinates -56, -18, 6).

Yet some theoretically-interesting regions of ISC emerge as well. Inter-subject activity in the cingulate gyrus, a component of the limbic system associated with emotion regulation (Ochsner & Gross, 2005), is significantly correlated across all experimental conditions (max r = 0.063, FDR-corrected p < 0.001, MNI-152 coordinates 0, -22, 30). Additionally, there is a region of significant ISC across all conditions in vmPFC (max r = 0.067, FDR-corrected p < 0.001, MNI-152 coordinates 0, 60, 0). Although in each case these correlations are relatively low in absolute terms, it is important to bear in mind that they reflect the average correlation across *all* experimental conditions – even those expected to

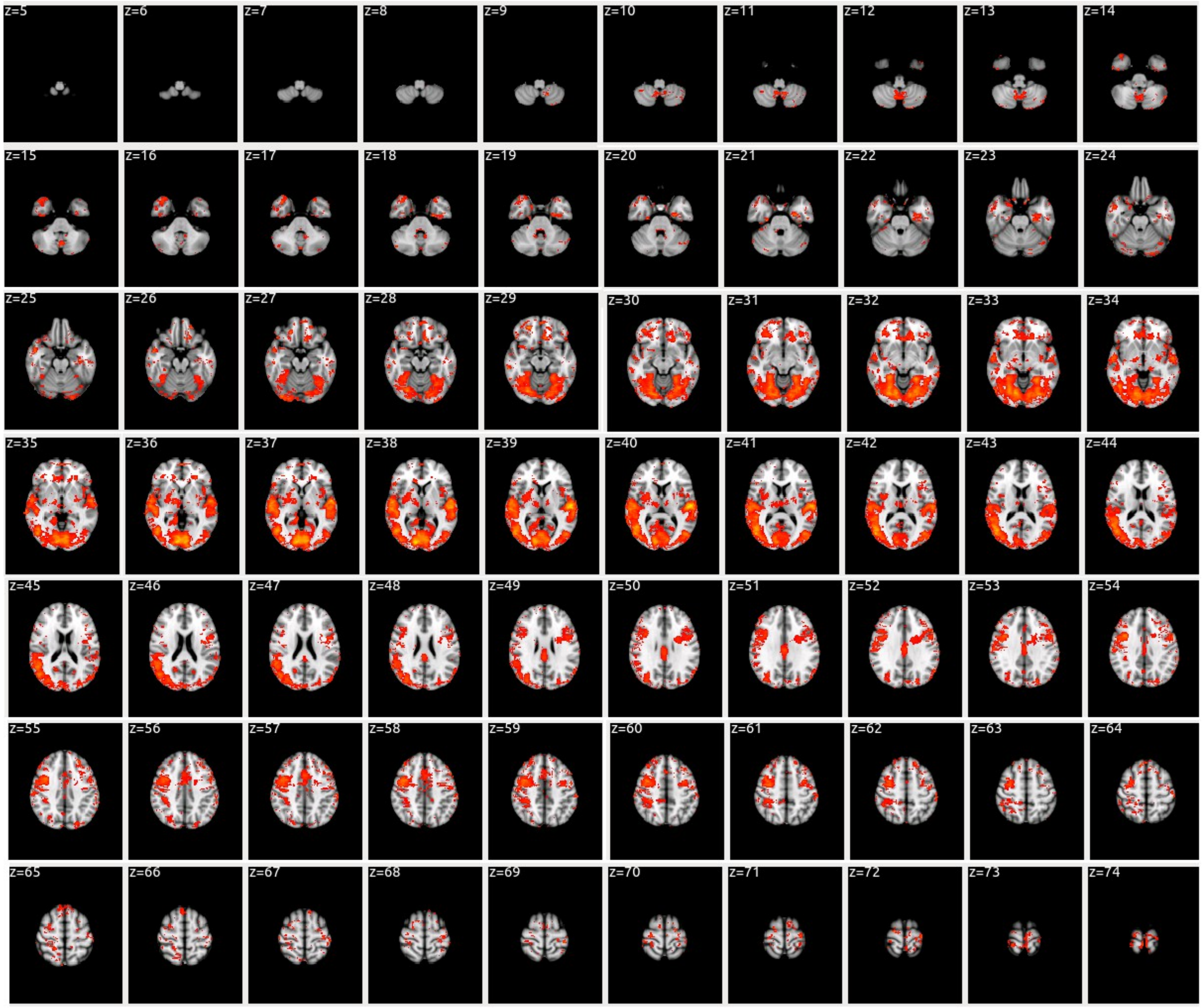

Figure 1. Axial image series showing areas of significant inter-subject correlation across all conditions, thresholded at $p < 0.05$, FDR corrected.

yield comparatively low ISC – and they are nonetheless significantly higher than ISC in other regions by even the most stringent standard (p<0.001).

However, no statistically-significant differences in whole-brain ISC maps were found between conditions using the procedure provided by the Matlab toolbox (sum ZPF = 25000; see Kauppi, Pajula, & Tohka, 2014). The ADT-driven differences in overall ISC reported by Weber et al. (2011) were therefore not directly replicated by this analysis, contrary to the prediction made in H1.

Nonetheless, the inferential statistics for ISC comparison are not yet fully understood, and it is important to note that this first analysis was an indirect replication – the whole-brain permutation test devised by Kauppi, Pajula, and Tohka (2014) differs substantially from the ROI-based statistical procedure used by Weber et al. (2011). Moreover, the motivation for the current study is not to look for whole-brain differences in ISC, but rather to see whether ISC in particular ROIs differs systematically in response to disposition-consistent or -inconsistent moral content. To more closely approximate the procedure of Weber et al. (2011), a ROI-driven approach was applied by computing subject-pairwise correlations in the mean time-series data from the pre-identified regions of interest, rather than conducting a voxel-wise whole-brain search for significant correlations.

First, analysis of variance (ANOVA) was used to assess how the broad ADT categories (disposition-consistent, disposition-inconsistent, or amoral control stimuli) influenced ISC in each ROI. Results are summarized in Tables 1a and 1b. Strongly significant differences were found for the left amygdala (F(2,3462) = 7.107, p = 0.001), right STS (F(2,3462) = 10.896, p < 0.001), right TPJ (F(2,3462) = 16.428, p < 0.001), and left TPJ

Table 1a

ANOVA for ISC by Disposition Level (Consistent/Inconsistent/Amoral)

ROI	F(2,3462)	Sig.
dmPFC ISC	.310	.734
lTPJ ISC	5.415*	.004*
mmPFC ISC	1.923	.146
PC ISC	1.262	.283
rSTS ISC	10.896*	.000*
rTPJ ISC	16.428*	.000*
vmPFC (Saxe) ISC	1.407	.245
lAMYG ISC	7.107*	.001*
lvmPFC (S&G, integrative>utilitarian) ISC	1.464	.232
lvmPFC (S&G, integrative>emotional) ISC	0.02	.985
rAMYG ISC	1.407	.245
rvmPFC (S&G, integrative>utilitarian) ISC	.067	.935
rvmPFC (S&G, integrative>emotional) ISC	.200	.819

* The F value is significant at the 0.01 level.

Table 1b

Post-hoc tests (Tukey HSD) for regions with a significant F value

	Disposition	Disposition	Mean Value		Mean Difference	Std.	
ROI	Level (I)	Level (J)	(I)	(J)	(I-J)	Err.	Sig.
lTPJ ISC	amoral	inconsistent	.176	.148	.028*	.009	.003*
		consistent	.176	.159	.017	.009	.135
	inconsistent	amoral	.148	.176	-.028*	.009	.003*
		consistent	.148	.159	-.012	.007	.220
	consistent	amoral	.159	.176	-.017	.009	.135
		inconsistent	.159	.147	.012	.007	.220
rSTS ISC	amoral	inconsistent	.225	.190	.034*	.009	.001*
		consistent	.225	.183	.042*	.009	.000*
	inconsistent	amoral	.190	.225	-.034*	.009	.001*
		consistent	.190	.183	.008	.008	.507
	consistent	amoral	.183	.225	-.042*	.009	.000*
		inconsistent	.183	.190	-.008	.008	.507
rTPJ ISC	amoral	inconsistent	.217	.201	.016	.009	.193
		consistent	.217	.243	-.026*	.009	.010*
	inconsistent	amoral	.201	.217	-.016	.009	.193
		consistent	.201	.243	-.042*	.007	.000*
	consistent	amoral	.243	.217	.026*	.009	.010*
		inconsistent	.243	.201	.042*	.007	.000*
lAMYG ISC	amoral	inconsistent	.029	.027	-.004	.008	.872
		consistent	.029	.049	-.025*	.008	.005*
	inconsistent	amoral	.027	.029	.004	.008	.872
		consistent	.027	.049	-.021*	.007	.004*
	consistent	amoral	.049	.029	.025*	.008	.005*
		inconsistent	.049	.027	.021*	.007	.004*

* The mean difference is significant at the 0.01 level.

(F(2,3462) = 5.415, p = 0.004). Contrary to H2b, no significant differences were found in vmPFC ISC. Because there are fewer amoral controls than stimulus videos, post-hoc analysis to determine the nature of these differences was conducted using the Tukey method, which is conservative when group sizes are unequal. As predicted in H2a, left amygdala ISC is significantly higher in disposition-consistent conditions compared to both disposition-inconsistent conditions (mean difference = 0.021, p = 0.004) and amoral controls (mean difference = 0.025, p = 0.005), although ISC does not significantly differ between disposition-inconsistent and amoral scenes (mean difference = 0.004, p = 0.872).

Consistent with H1, right TPJ ISC is also significantly higher in disposition-consistent conditions compared to disposition-inconsistent conditions (mean difference = 0.042, p < 0.001) as well as amoral controls (mean difference = 0.026, p = 0.010); rTPJ ISC is lowest in disposition-inconsistent scenes, although the difference between disposition-inconsistent and amoral scenes again falls short of statistical significance (mean difference = 0.015, p = .223).

Conversely, contrary to H1, left TPJ ISC is highest during amoral control scenes and lowest in disposition-inconsistent ones. The pairwise difference between amoral and disposition-inconsistent scenes reaches significance (mean difference = 0.028, p = 0.003), while the difference between amoral and disposition-consistent scenes falls short (mean difference = 0.016, p = 0.135). Disposition-consistent scenes also do not differ significantly from disposition-inconsistent ones (mean difference = 0.011, p = 0.220).

ISCs in right STS have a similar pattern to those in left TPJ: amoral scenes yield significantly higher ISC than disposition-consistent (mean difference = 0.042, p < 0.001) and disposition-inconsistent scenes (mean difference = 0.033, p = 0.001), but ISCs do not differ significantly between disposition-consistent and disposition-inconsistent scenes (mean difference = 0.008, p = 0.507).

A second ANOVA was conducted to distinguish between the valence of disposition-consistent or -inconsistent conduct (e.g. good-positive vs. bad-negative). These results implicate a wider array of brain regions. Mean ISCs in left amygdala (F(4,3460) = 3.960, p = 0.003), right STS (F(4,3460) = 10.010, p < 0.001), right TPJ (F(4,3460) = 52.296, p < 0.001) and left TPJ (F(4,3460) = 29.399, p < 0.001) again differ significantly across conditions, but so too do ISCs in dmPFC (F(4,3460) = 8.471, p < 0.001), mmPFC (F(4,3460) = 5.269, p < 0.001), and precuneus (F(4,3460) = 24.465, p < 0.001). Consistent with H1, in every case except for rSTS, greatest ISC is observed in the negatively-valenced disposition-consistent condition -- the punishment of wrongdoers generally maximizes ISC, consistent with the results of Weber et al. (2011). In particular, the negatively-valenced disposition-consistent condition (bad-negative) consistently produces higher ISC than its positively-valenced (good-positive) counterpart, with mean differences that are strongly significant, in every ROI *except* the amygdala. Although the general trend in amygdala ISC follows the expected pattern, there is no significant difference between the good-positive and bad-negative conditions (mean difference = 0.009, p = 0.867). Overall, these results largely support H1 but not H2. See Tables 2a and 2b for a full summary of mean differences by condition.

A final test was conducted to determine whether the strength of ISC varies by participants' self-reported trait empathy levels. None of the regions tested showed significant

Table 2a

ANOVA for ISC by Experimental Condition (good-positive, bad-negative, good-negative, bad-positive, amoral)

ROI	F(4,3460)	Sig.
dmPFC ISC	8.471*	.000*
lTPJ ISC	29.399*	.000*
mmPFC ISC	5.269*	.000*
PC ISC	24.465*	.000*
rSTS ISC	10.010*	.000*
rTPJ ISC	52.296*	.000*
vmPFC (Saxe) ISC	1.111	.349
lAMYG ISC	3.960*	.003*
lvmPFC (S&G, integrative>utilitarian) ISC	1.120	.345
lvmPFC (S&G, integrative>emotional) ISC	.212	.932
rAMYG ISC	.928	.447
rvmPFC (S&G, integrative>utilitarian) ISC	.781	.537
rvmPFC (S&G, integrative>emotional) ISC	.471	.757

* The F value is significant at the 0.01 level.

Table 2b

Post-hoc tests (Tukey HSD) for regions with a significant F value.

			Mean Value		Mean		
					Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
dmPFC ISC	good-positive	bad-negative	.026	.083	-.057*	.011	.000*
		good-negative	.026	.037	-.011	.011	.838
		bad-positive	.026	.063	-.037*	.011	.007*
		amoral	.026	.048	-.022	.011	.258
	bad-negative	good-positive	.083	.026	.057*	.011	.000*
		good-negative	.083	.037	.046*	.011	.000*
		bad-positive	.083	.063	.020	.011	.324
		amoral	.083	.048	.035*	.011	.010*
	good-negative	good-positive	.037	.026	.011	.011	.838
		bad-negative	.037	.083	-.046*	.011	.000*
		bad-positive	.037	.063	-.025	.011	.132
		amoral	.037	.048	-.011	.011	.865
	bad-positive	good-positive	.063	.026	.037*	.011	.007*
		bad-negative	.063	.083	-.020	.011	.324
		good-negative	.063	.037	.025	.011	.132
		amoral	.063	.048	.015	.011	.650
	amoral	good-positive	.048	.026	.022	.011	.258
		bad-negative	.048	.083	-.035*	.011	.010*
		good-negative	.048	.037	.011	.011	.865
		bad-positive	.048	.063	-.015	.011	0.65

			Mean Value		Mean		
					Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
lTPJ ISC	good-positive	bad-negative	.111	.208	-.097*	.010	.000*
		good-negative	.111	.133	-.022	.010	.167
		bad-positive	.111	.162	-.052*	.010	.000*
		amoral	.111	.176	-.065*	.010	.000*
	bad-negative	good-positive	.208	.111	.097*	.010	.000*
		good-negative	.208	.133	.075*	.010	.000*
		bad-positive	.208	.162	.046*	.010	.000*
		amoral	.208	.176	.032*	.010	.010*
	good-negative	good-positive	.133	.111	.022	.010	.167
		bad-negative	.133	.208	-.075*	.010	.000*
		bad-positive	.133	.162	-.030*	.010	.022*
		amoral	.133	.176	-.043*	.010	.000*
	bad-positive	good-positive	.162	.111	.052*	.010	.000*
		bad-negative	.162	.208	-.046*	.010	.000*
		good-negative	.162	.133	.030*	.010	.022*
		amoral	.162	.176	-.013	.010	.647
	amoral	good-positive	.176	.111	.065*	.010	.000*
		bad-negative	.176	.208	-.032*	.010	.010*
		good-negative	.176	.133	.043*	.010	.000*
		bad-positive	.176	.162	.013	.010	.647

			Mean Value		Mean		
					Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
mmPFC ISC	good-positive	bad-negative	.031	.075	-.043*	.011	.000*
		good-negative	.031	.046	-.015	.011	.612
		bad-positive	.031	.040	-.009	.011	.918
		amoral	.031	.036	-.005	.011	.993
	bad-negative	good-positive	.075	.031	.043*	.011	.000*
		good-negative	.075	.046	.028	.011	.057
		bad-positive	.075	.040	.034*	.011	.010*
		amoral	.075	.036	.039*	.011	.002*
	good-negative	good-positive	.046	.031	.015	.011	.612
		bad-negative	.046	.075	-.028	.011	.057
		bad-positive	.046	.040	.006	.011	.977
		amoral	.046	.036	.010	.011	.858
	bad-positive	good-positive	.040	.031	.009	.011	.918
		bad-negative	.040	.075	-.034*	.011	.010*
		good-negative	.040	.046	-.006	.011	.977
		amoral	.040	.036	.004	.011	.994
	amoral	good-positive	.036	.031	.005	.011	0.99
		bad-negative	.036	.075	-.039*	.011	.002*
		good-negative	.036	.046	-.010	.011	.858
		bad-positive	.036	.040	-.004	.011	.994

			Mean Value		Mean		
					Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
PC ISC	good-positive	bad-negative	.082	.180	-.098*	.010	.000*
		good-negative	.082	.139	-.056*	.010	.000*
		bad-positive	.082	.145	-.063*	.010	.000*
		amoral	.082	.131	-.049*	.010	.000*
	bad-negative	good-positive	.180	.082	.098*	.010	.000*
		good-negative	.180	.139	.041*	.010	.000*
		bad-positive	.180	.145	.035*	.010	.004*
		amoral	.180	.131	.048*	.010	.000*
	good-negative	good-positive	.139	.082	.056*	.010	.000*
		bad-negative	.139	.180	-.041*	.010	.000*
		bad-positive	.139	.145	-.006	.010	.972
		amoral	.139	.131	.007	.010	.956
	bad-positive	good-positive	.145	.082	.063*	.010	.000*
		bad-negative	.145	.180	-.035*	.010	.004*
		good-negative	.145	.139	.006	.010	.972
		amoral	.145	.131	.013	.010	.677
	amoral	good-positive	.131	.082	.049*	.010	.000*
		bad-negative	.131	.180	-.048*	.010	.000*
		good-negative	.131	.139	-.007	.010	.956
		bad-positive	.131	.145	-.013	.010	.677

			Mean Value		Mean Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
rSTS ISC	good-positive	bad-negative	.166	.199	-.033*	.011	.017*
		good-negative	.166	.175	-.009	.011	.906
		bad-positive	.166	.215	-.040*	.011	.001*
		amoral	.166	.217	-.059*	.011	.000*
	bad-negative	good-positive	.199	.166	.033*	.011	.017*
		good-negative	.199	.175	.024	.011	.171
		bad-positive	.199	.215	-.007	.011	.957
		amoral	.199	.217	-.026	.011	.106
	good-negative	good-positive	.175	.166	.009	.011	.906
		bad-negative	.175	.199	-.024	.011	.171
		bad-positive	.175	.215	-.031*	.011	.029*
		amoral	.175	.217	-.050*	.011	.000*
	bad-positive	good-positive	.215	.166	.040*	.011	.001*
		bad-negative	.215	.199	.007	.011	.957
		good-negative	.215	.175	.031*	.011	.029*
		amoral	.215	.217	-.018	.011	.411
	amoral	good-positive	.217	.166	.059*	.011	.000*
		bad-negative	.217	.199	.026	.011	.106
		good-negative	.217	.175	.050*	.011	.000*
		bad-positive	.217	.215	.018	.011	.411

			Mean Value		Mean		
					Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
rTPJ ISC	good-positive	bad-negative	.177	.309	-.132*	.010	.000*
		good-negative	177	.188	-.010	.010	.844
		bad-positive	177	.215	-.037*	.010	.002*
		amoral	177	.217	-.040*	.010	.001*
	bad-negative	good-positive	.309	177	.132*	.010	.000*
		good-negative	.309	.188	.121*	.010	.000*
		bad-positive	.309	.215	.095*	.010	.000*
		amoral	.309	.217	.092*	.010	.000*
	good-negative	good-positive	.188	177	.010	.010	.844
		bad-negative	.188	.309	-.121*	.010	.000*
		bad-positive	.188	.215	-.027	.010	.062
		amoral	.188	.217	-.029*	.010	.034*
	bad-positive	good-positive	.215	177	.037*	.010	.002*
		bad-negative	.215	.309	-.095*	.010	.000*
		good-negative	.215	.188	.027	.010	.062
		amoral	.215	.217	-.002	.010	1.000
	amoral	good-positive	.217	177	.040*	.010	.001*
		bad-negative	.217	.309	-.092*	.010	.000*
		good-negative	.217	.188	.029*	.010	.034*
		bad-positive	.217	.215	.002	.010	1.000

			Mean Value		Mean		
					Difference	Std.	
ROI	Condition (I)	Condition (J)	(I)	(J)	(I-J)	Error	Sig.
lAMYG ISC	good-positive	bad-negative	.044	.053	-.009	.009	.867
		good-negative	.044	.031	.013	.009	.647
		bad-positive	.044	.024	.021	.009	.183
		amoral	.044	.023	.021	.009	.175
	bad-negative	good-positive	.053	.044	.009	.009	.867
		good-negative	.053	.031	.022	.009	.131
		bad-positive	.053	.024	.030*	.009	.013*
		amoral	.053	.023	.030*	.009	.013*
	good-negative	good-positive	.031	.044	-.013	.009	.647
		bad-negative	.031	.053	-.022	.009	.131
		bad-positive	.031	.024	.008	.009	.924
		amoral	.031	.023	.008	.009	.917
	bad-positive	good-positive	.024	.044	-.021	.009	.183
		bad-negative	.024	.053	-.030*	.009	.013*
		good-negative	.024	.031	-.008	.009	.924
		amoral	.024	.023	.000	.009	1.000
	amoral	good-positive	.023	.044	-.021	.009	.175
		bad-negative	.023	.053	-.030*	.009	.013*
		good-negative	.023	.031	-.008	.009	.917
		bad-positive	.023	.024	.000	.009	1.000

* The mean difference is significant at the 0.05 level. All mean and difference values have been rounded.

differences in ISC among empathy-matched pairs compared to empathy-mismatched pairs, suggesting that differences in empathy do not drive differences in ISC.

PPI analysis. While the simple subtraction logic of brain mapping studies can offer insight into how the brain *segregates* information, functional connectivity analysis is interested in how the brain *integrates* information (Friston, 1994). Functional connectivity was evaluated using psychophysiological interaction (PPI) analysis. For each participant, an interaction regressor is created by taking the product of the time-course in a physiological seed region and the time-course of the psychological task, and this regressor is used as an explanatory variable in GLM analysis (O'Reilly et al., 2012). The resulting parameter estimates indicate which voxels are co-activated with the seed region under one experimental condition but not another.

In order to examine how connectivity among brain regions varies in response to narrative content, a multi-level PPI analysis was conducted through FEAT in accordance with standard guidelines provided in the FSL documentation. On the first level, data from all functional runs for each subject were modeled using three explanatory variables (EVs): a psychological regressor representing the experimental condition, a physiological regressor representing the time-series of BOLD signal within a particular ROI, and the interaction of physiological and psychological regressors (the PPI EV). This procedure was completed for each ROI mask (see above). To facilitate a finely-grained analysis of disposition-consistent content, separate analyses were used to distinguish between disposition-consistent contrasts: positive-valence (good-positive > good-negative) and negative-valence (bad-negative > bad-positive). Second-level analyses were conducted to combine each subject's data across all functional runs. Finally, third-level analyses assessed the PPI across subjects using a mixed-effects model (FLAME 1+2). All results reported here are family-wise error-rate corrected using cluster-extent thresholding with a Z-threshold of 2.3 and a cluster p-threshold of 0.05.

First, all participants were combined into an aggregate mean third-level image. There were no significant results for the left or right amygdala seed regions in either psychological contrast (good-positive > good-negative or bad-negative > bad-positive). Among the seed regions derived from Shenhav and Greene (2014), only the right vmPFC (5mm sphere around MNI-152 coordinates 2, 42, -14) yielded significant results. When contrasting bad-negative > bad-positive scenes, the right vmPFC exhibited greater functional connectivity in a region near the left temporal pole (cluster p = 0.039, max-Z = 3.43 at MNI-152 coordinates -30, 4, -34). The temporal pole is adjacent to and strongly interconnected with the amygdala, and meta-analysis indicates good evidence that the temporal pole plays a role in integrating perception and emotion (Olson, Plotzker, & Ezzyat, 2007).

However, H4 assumes that participants may vary in the extent to which the moral scenarios presented will elicit emotional responses. An additional set of exploratory third-level PPI analyses were conducted by dividing subject by self-reported trait empathy levels.

When using the right amygdala as a seed region, an interesting pattern emerges (Figure 2): the connectivity pattern for the empathy *high* > *low* contrast when *good* characters are *rewarded* is strikingly similar to the pattern for the empathy *low* > *high* contrast when *bad* characters are *punished*. When comparing positively-valenced content (good-positive scenes > good-negative scenes), the right amygdala shows greater connectivity with the inferior portion of parietal cortex on both the left (cluster p = 0.043, max-Z = 3.88 at MNI-152 coordinates -40, -36, 28) and right (cluster p = 0.002, max-Z = 4.14 at MNI-152 coordinates 24, -44, 30) in high-empathy individuals compared to low-empathy individuals. Yet when comparing negatively-valenced content (bad-negative scenes > bad-positive scenes), a very similar connectivity pattern emerges for *low*-empathy individuals compared to high-empathy ones (left: cluster p = 0.026, max-Z = 3.33 at MNI-152 coordinates -22, -18, 38; right: cluster p = 0.031, max-Z = 4.04 at MNI-152 coordinates 22, -36, 18).

The inferior parietal lobule was selectively activated with the same pattern of results when using the left vmPFC mask derived from Shenhav and Greene's (2014) integrative > utilitarian moral judgment contrast as the seed region (5mm sphere around MNI-152 coordinates -6, 26, -16). For the underlying psychological contrast good-positive > good-negative, high empathy participants show greater co-activation of left vmPFC with several regions in occipital, parietal, and frontal cortex when compared to low-empathy participants (see Table 3 for full results). The largest of these clusters overlaps substantially with the inferior parietal region identified by the amygdala-seed analysis (cluster p < 0.001, left-side max-Z = 4.29 at MNI-152 coordinates -36, -14, 26; right-side max-Z = 4.29 at MNI-152 coordinates 30, -34, 30). Conversely, when the underlying psychological contrast for the PPI is bad-negative > bad-positive, it is instead low-empathy participants who exhibit greater co-activation of vmPFC with the inferior parietal lobule when compared to high-empathy participants (cluster p = 0.008, max-Z = 3.37 at 36, -26, 22). The overlap in thresholded activation maps for the left vmPFC seed region is depicted in Figure 3.

Table 3

Activation Table for PPI Analysis (Shenhav & Greene ROIs)

Third-Level PPI Contrast	First-Level Psychological Contrast	Seed Region	Cluster Index	Voxels	Cluster p-value	Max Z	Coordinates (mm)
Mean (All Participants)	Good-positive > good-negative	rvmPFC (S&G, integrative > utilitarian)	1	280	0.044	3.20	(32, -66, 58)
	Bad-negative > bad-positive	rvmpFC (S&G, integrative > utilitarian)	1	319	0.039	3.42	(4, -34, -37)
Empathy High > Low	Good-positive > good-negative	rAMYG	1	402	0.043	3.88	(-40, -36, 28)
			2	691	0.002	4.14	(24, -44, 30)
		lvmPFC (S&G, integrative > utilitarian)	1	339	0.035	3.61	(-20, 28, 14)
			2	364	0.024	3.31	(12, -98, 8)
			3	604	< 0.001	3.52	(8, -76, 14)
			4	5181	< 0.001	4.29	(-36, -14, 26)
		lvmPFC (S&G, integrative > emotional)	1	289	0.036	3.41	(-8, 20, 66)
			2	311	0.024	3.34	(-40, -12, 54)
			3	346	0.013	3.88	(-56, -72, 18)
			4	819	< 0.001	3.62	(0, -46, 68)
			5	1057	< 0.001	3.89	(26, -6, 12)
		rvmPFC (S&G, integrative > utilitarian)	1	296	0.031	3.86	(-58, -2, -12)
			2	320	0.020	4.00	(12, -16, -14)
			3	333	0.016	3.37	(62, -6, 12)

			4	403	0.005	4.18	(-10, 50, 14)
			5	620	< 0.001	3.52	(-8, 24, 2)
		rvmPFC (S&G, integrative > emotional)	1	375	0.008	3.71	(-54, -70, 18)
			2	499	0.001	3.45	(10, -28, -4)
			3	640	< 0.001	3.75	(0, -46, 68)
			4	1401	< 0.001	4.05	(26, -8, 4)
Empathy Low > High	Bad-negative > bad-positive	rAMYG	1	398	0.031	4.04	(22, -36, 18)
			2	411	0.026	3.33	(-22, -18, 38)
		lvmPFC (S&G, integrative > utilitarian)	1	477	0.008	3.73	(36, -26, 22)

Combinations of contrasts and seed regions not listed here yield no significant clusters.

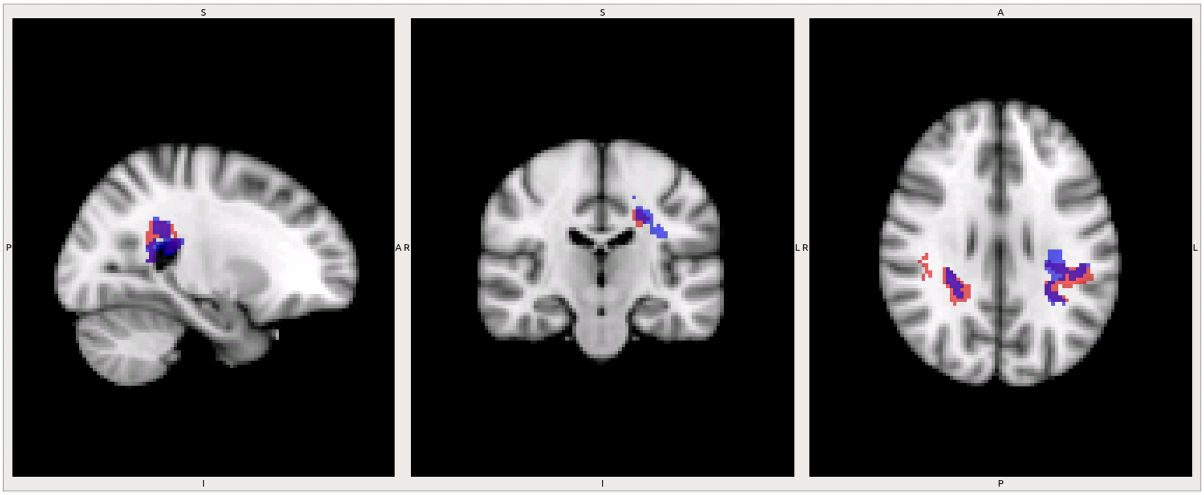

Figure 2. Overlap in PPI results with right amygdala as the seed region. Red is the group difference for the empathy high > low contrast with good-positive > good-negative as the underlying psychological regressor; blue is the group difference for empathy low > high with bad-negative > bad-positive as the underlying psychological regressor.

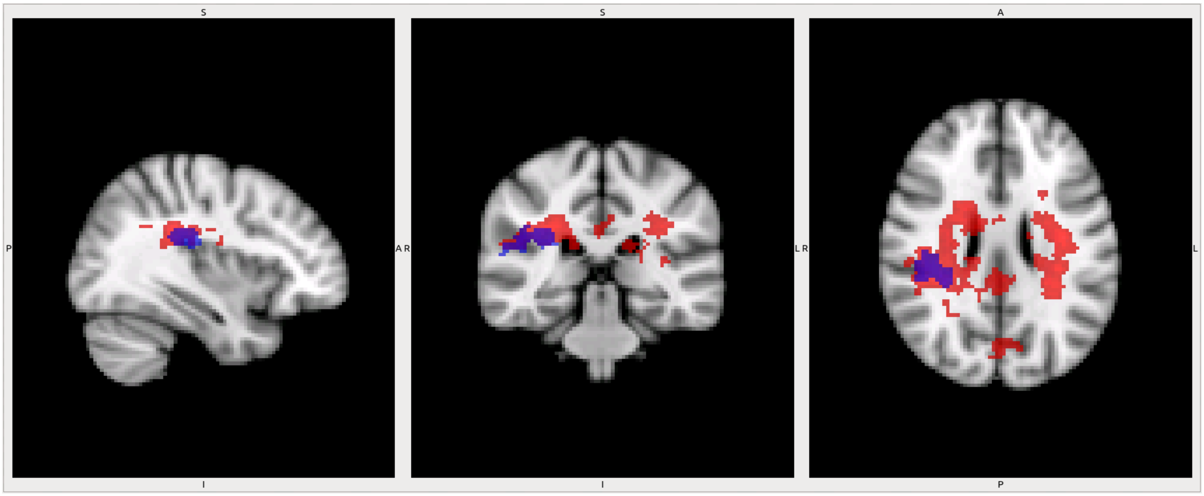

Figure 3. Overlap in PPI results with left vmPFC as the seed region. Red is the group difference for the empathy high > low contrast with good-positive > good-negative as the underlying psychological regressor; blue is the group difference for empathy low > high with bad-negative > bad-positive as the underlying psychological regressor.

Exploratory PPI analyses were also conducted using the ToM ROIs as seeds. Results for these seed regions are in Table 4. The mean PPI across all subjects in mmPFC, PC, and bilateral TPJ displays an interesting difference between the psychological contrasts. For the positively-valenced (good-positive > good-negative) contrast, these regions yield either no significant clusters in the mean PPI, or a few spatially-compact clusters. On the other hand, in each case, the negatively-valenced (bad-negative > bad-positive) contrast yields a substantially larger and more diverse connectivity pattern. See Figures 4 and 5 for examples. This trend suggests that negatively-valenced content modulates the coactivation of these seed ROIs with other regions more than positively-valenced content.

Effect of connectivity on affective disposition. Within subject coactivation of amygdala and vmPFC is not significantly predictive of participants' ADT index, which measures the participants' perception of disposition-consistent or -inconsistent content. In a linear regression model with ADT rating as the DV and coactivation between the amygdala and vmPFC seed regions as predictors, no variance in ADT rating is explained (R = 0.182, adjusted R^2 = -0.006). In an ANOVA, none of these coactivations vary significantly across experimental condition, although coactivation of left amygdala and vmPFC (5mm sphere around MNI-152 coordinates 0, 36, -10) comes the closest (F(4,325) = 2.162, p = 0.073; mean difference between bad-negative and control condition = 0.099, p = 0.105). A follow-up regression model using self-reported enjoyment as the dependent variable also yielded no meaningful predictive power (R = 0.172, adjusted R^2 = -0.009).

Table 4

Activation Table for PPI Analysis (Saxelab ToM ROIs)

Third-Level Contrast	First-Level Contrast	Seed ROI	Cluster Index	Voxels	Cluster p-value	Max Z	Coordinates (mm)
Mean (All Participants)	Good-positive > good-negative	PC	1	510	0.044	3.12	(34, -64, -48)
			2	703	0.008	3.38	(14, 6, 16)
		dmPFC	1	606	0.013	3.42	(-50, -8, -12)
			2	1238	< 0.001	3.42	(60, -18, -10)
		rTPJ	1	524	0.033	3.51	(8, -6, 76)
			2	755	0.004	3.65	(42, -36, 14)
			3	919	0.001	3.51	(42, -82, -4)
		rSTS	1	814	0.006	3.31	(52, 0, 22)
			2	7862	< 0.001	3.95	(34, -54, 14)
	Bad-negative > bad-positive	PC	1	588	0.039	3.82	(24, 38, -22)
			2	621	0.030	3.89	(-40, 32, 10)
			3	667	0.021	3.50	(-40, 6, 26)
			4	1254	< 0.001	3.48	(8, 20, 22)
			5	1479	< 0.001	3.59	(38, 32, 18)
			6	1989	< 0.001	3.88	(52, -66, -10)
			7	19954	< 0.001	4.66	(-48, -62, -20)
		dmPFC	1	501	0.038	3.78	(38, 48, 24)
			2	634	0.011	3.80	(-34, -86, -10)
			3	860	0.001	3.45	(64, -56, -14)
			4	1757	< 0.001	3.98	(-58, -32, 40)
		mmPFC	1	501	0.025	3.59	(-40, 4, 18)
			2	511	0.022	3.30	(38, 44, 24)
			3	517	0.021	3.25	(56, 8, 34)
			4	673	0.004	3.91	(-34, 46, 20)
			5	888	0.001	3.66	(-50, -36, -22)
			6	1021	< 0.001	3.73	(6, -70, -50)
			7	1197	< 0.001	4.36	(34, 0, 60)
			8	1719	< 0.001	4.04	(-58, -32, 40)
			9	5111	< 0.001	4.11	(52, -58, -28)
			10	5985	< 0.001	4.10	(30, -46, 64)
		rTPJ	1	526	0.039	4.12	(8, -38, 42)
			2	534	0.036	3.86	(66, -32, 34)

			3	1370	< 0.001	4.00	(-38, 4, 22)
			4	2360	< 0.001	4.15	(40, 44, 24)
			5	2687	< 0.001	4.25	(-28, 32, 30)
			6	5395	< 0.001	4.14	(6, 30, 42)
			7	8727	< 0.001	4.58	(-54. -60, -20)
		lTPJ	1	520	0.039	3.31	(38, 38, 30)
			2	706	0.007	3.41	(-30, -58, 60)
			3	1050	< 0.001	3.77	(-12, -70, -6)
			4	1684	< 0.001	3.86	(-36, -32, -22)
Empathy High > Low	Good-positive > good-negative	PC	1	668	0.009	3.44	(38, -12, -12)
		dmPFC	1	530	0.024	3.49	(8, -72, -6)
			2	592	0.013	3.35	(46, 4, -2)
		mmPFC	1	719	0.002	3.82	(32, -26, 24)
		vmPFC (Saxe)	1	798	0.001	3.64	(-4, -40, 20)
			2	3426	< 0.001	4.21	(-10, -24, -36)
		rSTS	1	1792	< 0.001	4.11	(-2, -30, 54)
			2	1823	< 0.001	4.10	(56, -50, 18)
			3	3135	< 0.001	3.92	(18, -54, 0)
Empathy Low > High	Bad-negative > bad-positive	PC	1	793	0.006	3.75	(-16, -48, -30)
			2	1353	< 0.001	3.74	(12, -74, -4)
			3	1361	< 0.001	3.60	(30, -12, 0)
			4	2036	< 0.001	4.09	(-18, -4, -12)
		mmPFC	1	461	0.035	3.77	(-16, 4, -10)
		vmPFC (Saxe)	1	854	< 0.001	3.50	(-34, -8, 62)
		rSTS	1	886	0.004	4.02	(4, -74, 10)
		rTPJ	1	1323	< 0.001	3.63	(10, -74, -4)
			2	1498	< 0.001	4.39	(-22, -4, -12)
		lTPJ	1	607	0.017	3.67	(8, -76, -4)

Combinations of contrasts and seed regions not listed here yield no significant clusters.

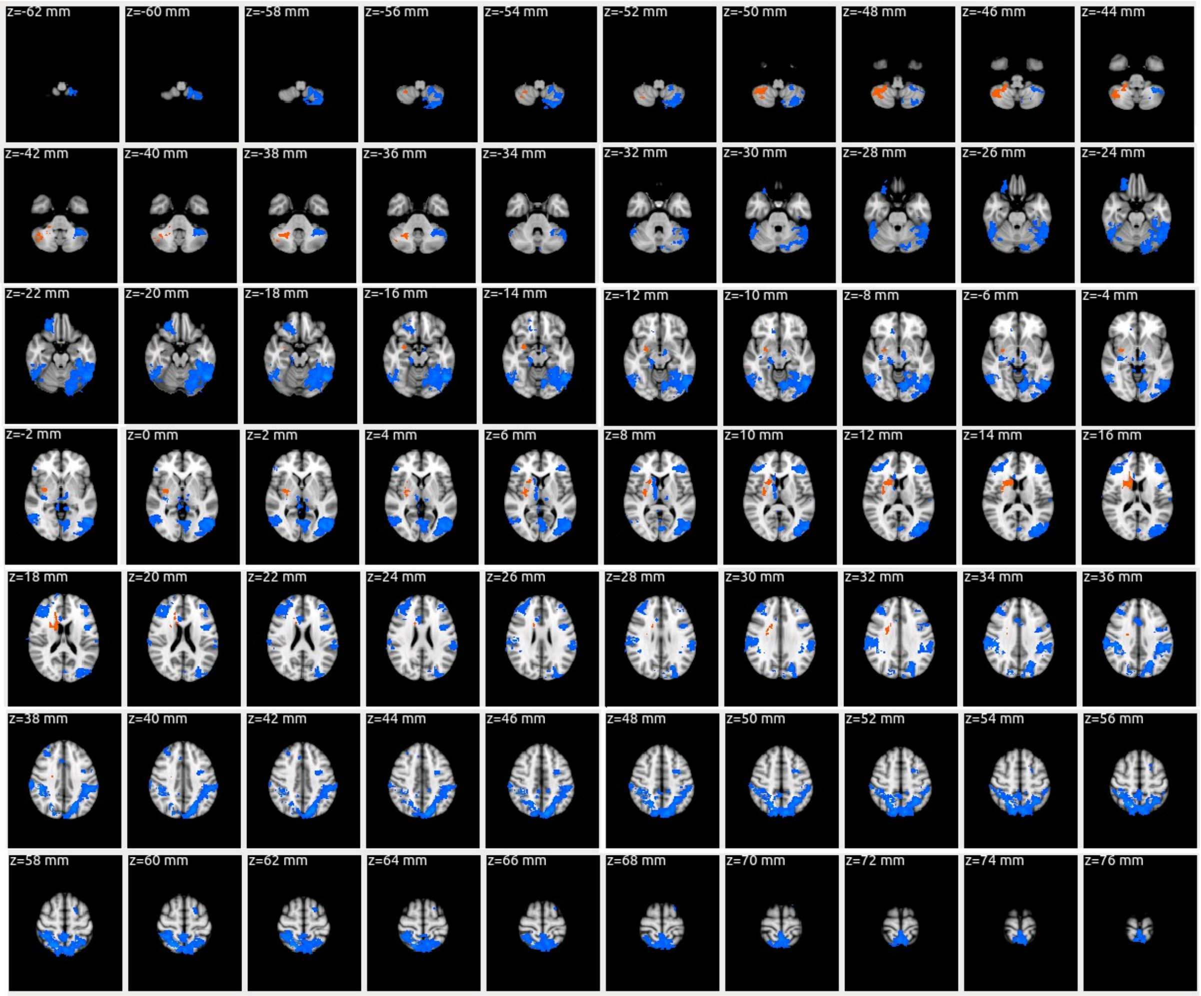

Figure 4. Mean PPI results (all participants) when using PC as a seed region; the positively-valenced contrast (good-positive > good-negative, in red) yields more limited differences in connectivity than the negatively-valenced contrast (bad-negative > bad-positive, in blue).

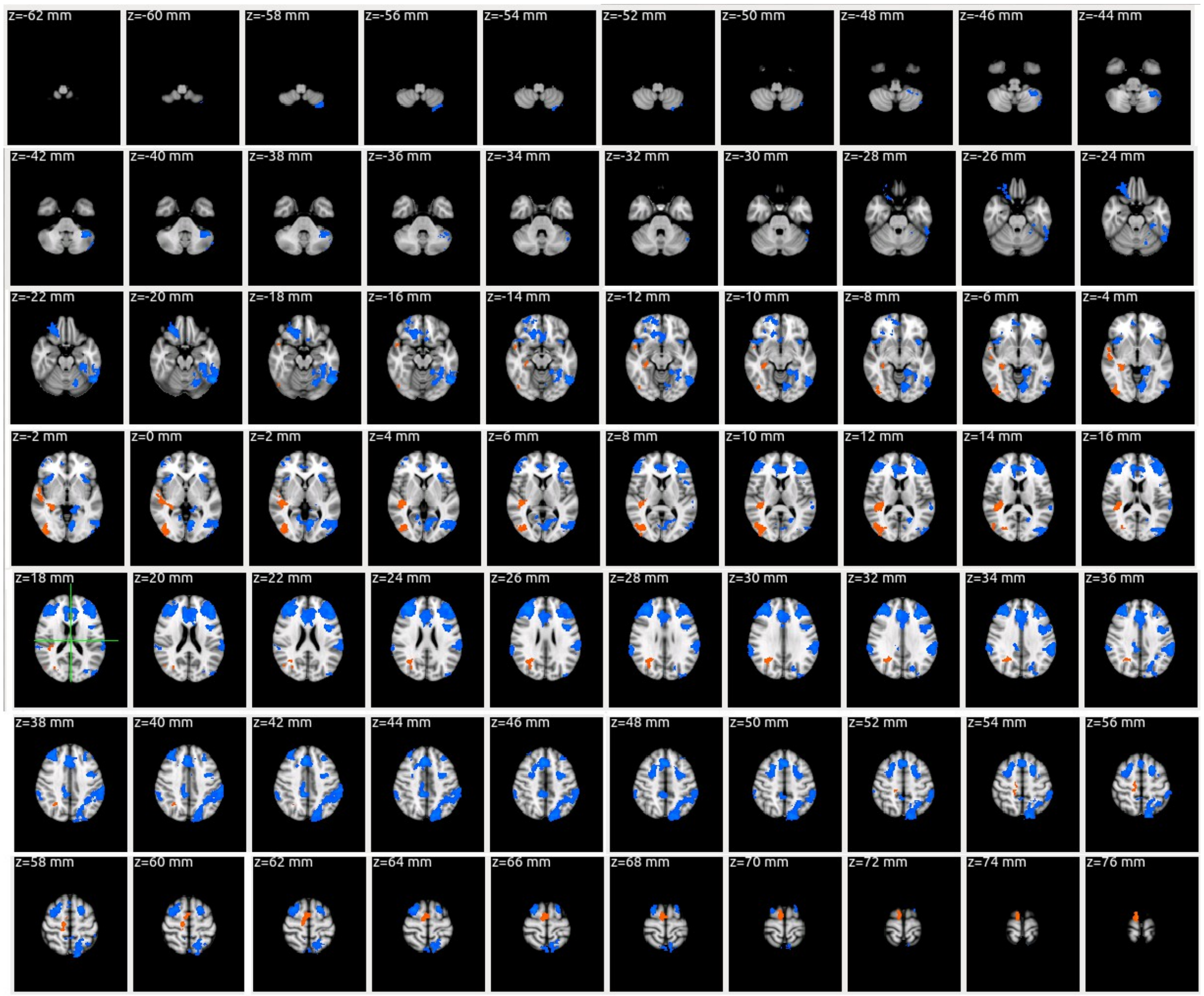

Figure 5. Mean PPI results (all participants) when using rTPJ as a seed region; the positively-valenced contrast (good-positive > good-negative, in red) yields more limited differences in connectivity than the negatively-valenced contrast (bad-negative > bad-positive, in blue).

Subsequent analysis revealed that when using enjoyment as the DV, model fit can be somewhat improved by only examining high-empathy participants (R = 0.349, adjusted R^2 = 0.045). In this exploratory model, within-subject coactivation of right amygdala and the vmPFC region identified by the Saxelab has a *negative* effect on enjoyment (standardized β = -0.361, t = -2.007, p = 0.047), while coactivation of left amygdala and vmPFC has a slightly weaker *positive* effect on enjoyment (standardized β = 0.334, t = 1.883 p = 0.062). When examining only low-empathy participants, a linear regression model again fails to explain variation in enjoyment (R = 0.183, adjusted R^2 = -0.043). In sum, these results fail to support the prediction made in H4.

Face-tracking Analysis

Data-collection pipeline. A facial movement time-series was extracted for each stimulus video from the recordings of a participant's think-aloud session. Face recognition and tracking was performed using the Clmtrackr library (https://github.com/auduno/clmtrackr) which uses the constrained local model technique for tracking facial features (Saragih, Lucey, & Cohn, 2009). The output from Clmtrackr is an array of 71 coordinate pairs identifying the location of specific facial landmarks. It also includes an SVM classifier that uses these features to provide automatic emotion detection in four dimensions (happy, sad, angry, surprised). The models provided by Clmtrackr were trained using sample images from the MUCT face database (http://www.milbo.org/muct/).

Because Clmtrackr is implemented in JavaScript, a somewhat unusual processing pipeline was necessary for automatic extraction of the facial time-series. First, the recordings for each participant were trimmed such that each think-aloud video begins when the stimulus video appears on screen for the participant and continues for 3 minutes (i.e., the length of the stimulus video). The trimmed recordings were then converted to the Theora format to facilitate use of the HTML5 ‘video’ element. Trimming and re-encoding were carried out using Mencoder (http://www.mplayerhq.hu) and Ffmpeg (http://www.ffmpeg.org) respectively, and both processes were automated with Python.

Next, a template page was created using HTML5 and JavaScript to run Clmtrackr on an arbitrary video. A Python script was then used to convert the template page into a series of unique HTML files - one for each think-aloud video. When opened in a Web browser, the HTML file for a particular video will automatically play that video and record the Clmtrackr output. The position and emotion parameters are updated every 500ms and the resultant time-series is packed into a JavaScript object. When the video is complete, a callback function submits the resultant data in JavaScript object notation (JSON) to an ad-hoc local HTTP server, and a Python common gateway interface (CGI) script saves the data for later analysis. The process of starting the ad-hoc server, opening each video, waiting for output, and then closing the browser was also automated in Python. All videos were processed in the Chromium Web browser, an open-source variant of Google Chrome for Linux. See Appendix A for source code for this pipeline.

Preprocessing. This procedure generated a substantial amount of raw data: each time-point consists of a 71x2x2 array of facial landmark coordinates in addition to the 1x4 emotion classifier output, and the output file for a single video contains more than 350 time-points, yielding nearly 400MB of data in total. In order to consolidate this information, several facial features of interest were computed, such as brow-to-brow distance, mouth width, eye height, and nose-to-brow distance. These features, in addition to the emotion classifier output, are the variables used in subsequent analyses. See Table 5 and Figure 6 for a full description of each feature and its underlying reference points.

Facial feature time-series were pre-processed using the Pandas data analysis library (http://pandas.pydata.org). First, the raw time-series data was read from the JSON file for each video. To account for minor variations in the length of the raw data, all series were trimmed to contain exactly 320 time-points (160s). Then, the facial features described above were computed for each time-point. The resultant time-series was smoothed using a 2s moving average and missing time-points were estimated using linear interpolation. To account for differences in magnitude across features, the time-series for each feature was standardized. A plot of a representative time-series from the emotion classifier is provided in Figure 7.

Statistical tests. Mirroring the procedure used for brain ISC above, subject-pairwise correlation in the time-series of each facial feature was computed by video. This data was first evaluated using ANOVA to compare mean facial correlation across ADT level (disposition-consistent, disposition-inconsistent, or amoral). Post-hoc tests of mean differences were computed using the Tukey method. The subject-pairwise correlation in mean emotion level (the average of all emotion scores in the Clmtrackr classifier) differs significantly between disposition levels ($F(2,3462) = 9.360$, $p < 0.001$). Pairwise correlation in mean emotion is greater

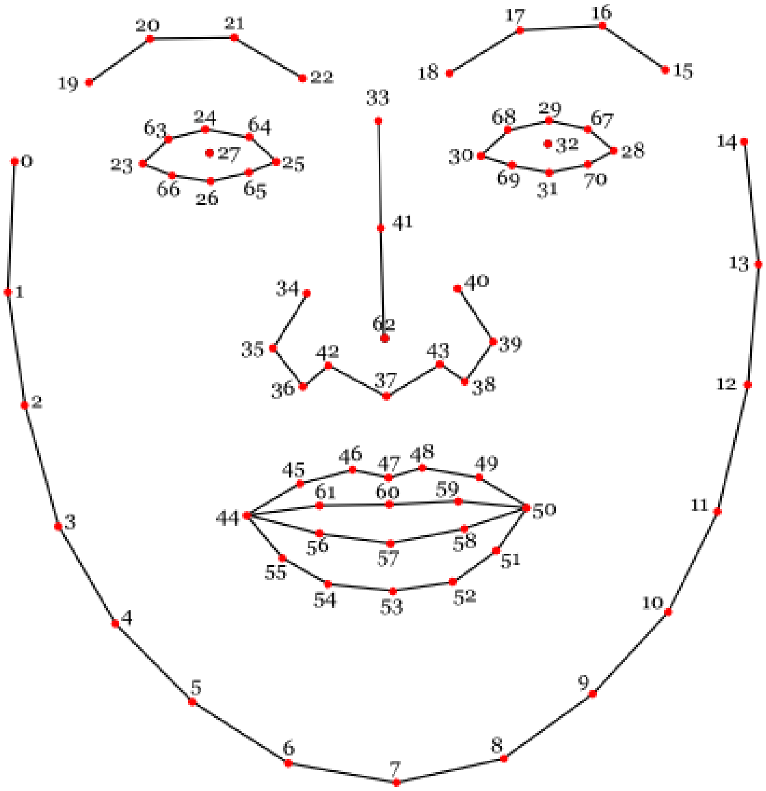

Figure 6. Facial anchor points tracked by clmtracker (from http://github.com/audunо/clmtrackr).

Table 5

Definition of computed facial features where m *is a function that takes the mean of its operands,* d *is a function that takes the Euclidean distance of its operands, and integer values are anchor point indices (see Figure 6).*

Feature	Calculation
Brow-to-Brow Distance (from center)	$d(20, 16)$
Brow-to-Brow Distance (from edge)	$d(22, 18)$
Mouth-to-nose distance	$m(d(35,44), d(39,50))$
Mouth height	$d(47, 53)$
Mouth width	$d(44, 50)$
Nose-to-brow distance	$d(33, 62)$
Eye height	$m(d(24, 26), d(29, 31))$
Eye width	$m(d(23, 25), d(28, 30))$

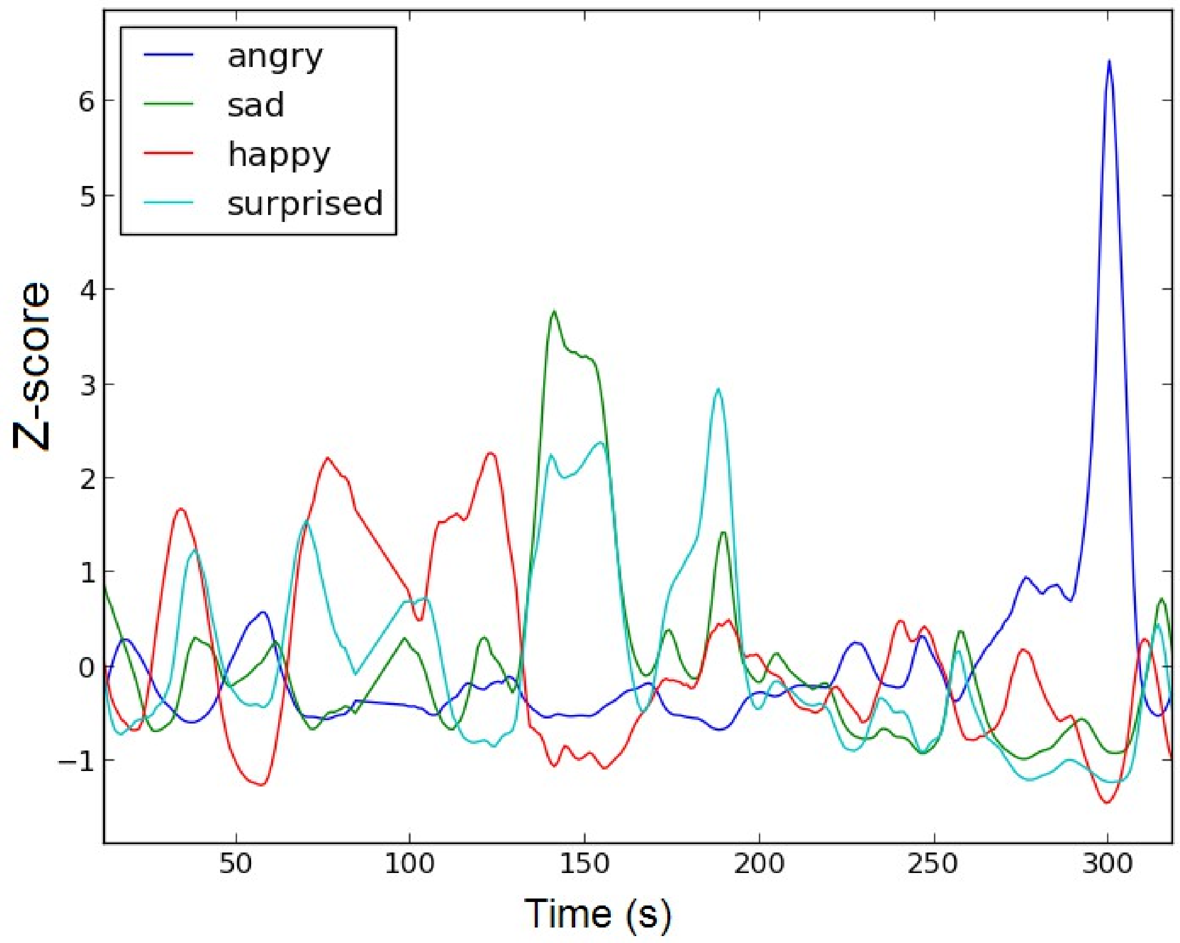

Figure 7. Sample emotion classifier time-series for one subject during one scene.

in disposition-inconsistent conditions than disposition-consistent ones (mean difference = 0.022, $p = 0.019$); in fact, while disposition-inconsistent conditions produce significantly higher facial correlations than amoral controls (mean difference = 0.043, $p < 0.001$), the difference between disposition-consistent conditions and amoral controls is not significant (mean difference = 0.020, $p = 0.116$). Likewise, subject-pairwise correlation in sadness level (as classified by Clmtrackr) displays the same pattern ($F(2,3462) = 4.997$, $p = 0.007$), with higher correlations in disposition-inconsistent conditions than disposition-consistent ones (mean difference = 0.018, $p = 0.077$) and amoral controls (mean difference = 0.031, $p = 0.008$). The anger dimension of the emotion classifier also displays the same trend but does not reach statistical significance ($F(2,3462) = 2.215$, $p = 0.109$). It seems that, as predicted by ADT, subjects react negatively to disposition-inconsistent outcomes; it is negative emotions (sadness, anger) which show condition-wise differences in between-subject facial correlations. Conversely, no corresponding increase in positive emotion (happiness) was found in disposition-consistent conditions. None of the lower-level facial features measured were found to be significant in this test, although eye height (how "open" the eyes are) was near the threshold ($F(2,3462) = 2.827$, $p = 0.059$). See Tables 6a and 6b for full results.

To investigate whether subject pairs differ systematically in the similarity of their facial expressions, a second ANOVA was conducted to determine if subject-pairwise correlations in facial expression vary by empathy level. Pairwise correlation in the mean emotion time-series differs significantly in this analysis ($F(2,3462) = 104.037$, $p < 0.001$) with higher correlations in pairs where both participants are high-empathy compared to pairs where both participants are low-empathy (mean difference = 0.151, $p < 0.001$) or where

Table 6a

ANOVA for mean pairwise facial expression correlation by disposition level (disposition-consistent, disposition-inconsistent, amoral)

Feature	F(2,3462)	Sig.
Anger	2.215	.109
Brow-to-Brow Distance (from center)	.383	.682
Brow-to-Brow Distance (from edge)	.191	.826
Eye height	2.827	.059
Eye width	.601	.548
Happiness	1.285	.277
Mean emotion	9.360*	.000*
Mouth-to-nose distance	.138	.871
Mouth height	1.862	.156
Mouth width	.709	.492
Nose-to-brow distance	1.053	.349
Sadness	4.997*	.007*
Surprise	1.954	.142

* The F-test is significant at the 0.01 level.

Table 6b

Post-hoc tests (Tukey HSD) for classified emotional features.

Feature	Disposition Level (I)	Disposition Level (J)	Mean Value (I)	(J)	Mean Difference (I-J)	Std. Err.	Sig.
Anger	amoral	inconsistent	.146	.145	.001	.009	.989
		consistent	.146	.131	.016	.009	.221
	inconsistent	amoral	.145	.146	-.001	.009	.989
		consistent	.145	.131	.014	.008	.151
	consistent	amoral	.131	.146	-.016	.009	.221
		inconsistent	.131	.145	-.014	.008	.151
Happiness	amoral	inconsistent	.130	.144	-.015	.009	.264
		consistent	.130	.142	-.013	.009	.384
	inconsistent	amoral	.144	.130	.015	.009	.264
		consistent	.144	.142	.002	.008	.954
	consistent	amoral	.142	.130	.013	.009	.384
		inconsistent	.142	.144	-.002	.008	.954
Mean Emotion	amoral	inconsistent	.362	.405	-.043*	.010	.000*
		consistent	.362	.382	-.020	.010	.116
	inconsistent	amoral	.405	.362	.043*	.010	.000*
		consistent	.405	.382	.022*	.008	.019*
	consistent	amoral	.382	.362	.020	.010	.116
		inconsistent	.382	.405	-.022*	.008	.019*
Sadness	amoral	inconsistent	.175	.206	-.031*	.010	.008*
		consistent	.175	.188	-.013	.010	.443
	inconsistent	amoral	.206	.175	.031*	.010	.008
		consistent	.206	.188	.018	.008	.077
	consistent	amoral	.188	.175	.013	.010	.443
		inconsistent	.188	.206	-.018	.008	.077
Surprise	amoral	inconsistent	.194	.212	-.018	.012	.266
		consistent	.194	.195	-.001	.012	.991
	inconsistent	amoral	.212	.194	.018	.012	.266
		consistent	.212	.195	.016	.009	.187
	consistent	amoral	.195	.194	.001	.012	.991
		inconsistent	.195	.212	-.016	.009	.187

* The mean difference is significant at the 0.05 level.

participants are mis-matched in empathy level (mean difference = 0.076, $p < 0.001$). Correlations in the level of classified sadness ($F(2,3462) = 41.052$, $p < 0.001$) and surprise ($F(2,3462) = 126.072$, $p < 0.001$) obey the same trend, with the highest subject-pairwise correlations when both subjects are more empathic and the lowest subject-pairwise correlations when both subjects are less empathic (all mean differences between high/low/mismatched groups are significant, $p < 0.001$). Correlations in anger and happiness levels were not significantly different in this model, although happiness displays the same trend at a near-significant level ($F(2,3462) = 2.775$, $p = 0.062$; both-high vs. both-low mean difference = 0.023, $p = 0.051$).

A linear regression analysis was also conducted to examine the relationship between intersubject correlations in the brain and in the face. Although intersubject correlations in both brain and face differ systematically by condition, they do not correlate strongly with each other. Contrary to H3, neuronal ISC predicts only a very small amount of correlation in facial mean emotion level ($R = 0.101$, adjusted $R^2 = 0.007$). None of the vmPFC seed regions are significant predictors in this model. The amygdala is weakly predictive and does not reach significance (standardized $\beta = 0.036$, $t = 1.696$, $p = 0.090$). Several of Saxe's ToM regions are weak *negative* predictors: mmPFC (standardized $\beta = -0.043$, $t = 1.967$, $p = 0.049$), right STS (standardized $\beta = -0.061$, $t = -2.829$, $p = 0.005$), and right TPJ (standardized $\beta = -0.065$, $t = -2.699$, $p = 0.007$). The only significant positively-signed predictor was left TPJ (standardized $\beta = 0.061$, $t = 2.451$, $p = 0.014$). In other words, generally speaking, the more similar subject-pairs are in the time-course of their facial displays of emotion, the less similar they are in the time-course of coactivation in these ROIs.

However, given the extremely low amount of overall variance explained by this model, that trend must be interpreted very cautiously. Collectively, the findings do not bear out the prediction made in H3.

Discussion

The hypothesized effects receive mixed support in this study. In keeping with the finding of Weber et al. (2011), ISCs in many of the morally-relevant ROIs examined here are higher in disposition-consistent conditions and highest when immoral characters are punished. The greater correlations in brain activity under these conditions supports the view that disposition-consistent moral content operates on commonly-held intuitive preferences, with individuals exhibiting a particular shared sensitivity to the punishment of norm-violators. Conversely, disposition-inconsistent content shows greater between-subject variability in brain activity, suggesting that individuals differ in how they process these counter-intuitive outcomes. Under this interpretation, the especially strong ISC when immoral characters are punished stems from the importance of altruistic punishment as a mechanism for encouraging norm-adherence; correcting the undesirable behavior of norm-violators improves collective outcomes more than encouraging the desirable behavior of norm-adherents. A strong intuition that wrongdoers must be punished is crucial to cooperation in human social groups, particularly in modern societies where group sizes are large and social ties are comparatively weak (Fehr & Gachter, 2002).

However, while inter-subject correlations in facial expression do vary systematically between disposition-consistent and disposition-inconsistent conditions, they do not match the pattern of neuronal ISC. Instead, facial displays of emotion are most similar across participants during disposition-inconsistent conditions. Consistent with ADT, these expectation-violating scenes produce a negatively-valenced emotional reaction which is reflected in participants' faces. However, previous research has demonstrated that facial expressions are not simply a neutral window into one's emotional state, but a means to communicate emotional information to others (Firth, 2009). For example, when observing someone else in pain, individuals empathically mirror pained facial expressions, but this mirroring is exaggerated when the empathizer is observed by the person in pain (Bavelas, Black, Lemery, & Mullet, 1986). Moreover, meta-analysis has shown that the facial expressions theorized by the widely-studied affect program theory (Eckman, 1993) generally are not reliably elicited by the mere experience of a particular emotion (Reisenzein, Studtmann, & Horstmann, 2013). Rather than directly conveying an emotional state, facial expressions may instead serve primarily as a means of conveying one's motivations to others (Parkinson, 2005). Drawing on this view, one explanation for the findings in this study is that facial expressions serve as a way for people to convey their rejection of morally-unacceptable disposition-inconsistent outcomes (e.g. rewarding a wrongdoer) moreso than to convey their acceptance of morally-laudable outcomes, which may be less communicatively salient because it is taken for granted. It is important to note that participants in this study were seated next to a research assistant during collection of the think-aloud data used for facial analysis. Further research might be conducted to determine if these emotional displays in response to moral narrative content are modulated by the presence of others. It may be, for example, that negative facial expressions signal the rejection of inequitable outcomes to others and thereby encourage punishment of norm-violators.

While the functional connectivity analyses did not support the hypothesized relationship between amygdala-vmPFC connectivity and affective disposition, they nonetheless reveal intriguing patterns to guide future research. The results indicate a complex systematic pattern of functional connectivity for both amygdala and vmPFC seed regions. Both regions exhibit significantly higher connectivity with the inferior parietal lobule during disposition-consistent conditions, but this relationship is moderated by trait empathy: the pattern of connectivity is the same for high-empathic individuals watching positively-valenced disposition-consistent content as it is for low-empathic individuals watching negatively-valenced disposition-consistent content. Although the inferior parietal lobule was not an *a priori* ROI, it forms the upper boundary of the TPJ and has been identified in other research as playing a role in automatic emotional processing (Lichev et al,, 2015; Radua et al., 2010).

Speculatively, if these connectivity patterns are taken to reflect the differential activity of an emotional judgment network, then a plausible explanation might be that more empathic people more readily share in the joy of others (while judging wrongdoers less harshly), whereas less empathic people find particular affective satisfaction in the punishment of wrongdoers (while taking no particular pleasure in the joy of others). However, empirical support for this idea remains sparse, since most studies linking morality with empathy generally examine how different moral scenarios elicit varying levels of empathic concern (e.g. Decety, Michalska, & Kinzler, 2012), or, alternatively, examine how trait empathy influences the propensity for moral judgment (e.g., Detert, Trevino, & Sweitzer, 2008). In building a philosophical argument, Prinz (2011) draws on altruistic punishment research to advance the claim that negatively-valenced affect is more motivating than empathy per se. Yet to my knowledge, no study has yet empirically studied the relationship between trait empathy and differing valences of moral content. The striking similarity of low-empathy participants during negatively-valenced conditions with high-empathy participants during positively-valenced conditions indicates that individual differences can moderate the effects of morally-salient content on brain activity, even though, in the aggregate, disposition-consistent outcomes increase ISC regardless of empathy level. The findings in this study support the idea that patterns of functional connectivity vary considerably between high- and low-empathy participants and thereby suggest that trait empathy is an important covariate for future research.

Additionally, the exploratory PPI analyses using ToM seed regions support the notion that negatively-valenced disposition-consistent content tends to recruit a more diverse network of brain regions than positively-valenced disposition-consistent content. An illustrative example is functional connectivity of the rTPJ in this data. In the positively-valenced contrast (good-positive > good-negative), PPI analysis yields 3 relatively small clusters in occiptal, temporal, and precentral frontal cortex. On the other hand, in the negatively-valenced contrast (bad-negative > bad-positive), there are 7 large clusters with uniformly-higher max-Z values spanning occipital, temporal, parietal, and frontal cortex. Of particular interest in the negatively-valenced condition is the apparent coactivation of rTPJ with brain regions widely-associated with executive decision-making such as the anterior cingulate cortex (ACC) and bilateral frontal pole. Likewise, using the precuneus as a seed region, the positively-valenced contrast yields 2 small clusters, whereas the negatively-valenced contrast yields 7 wider-reaching areas of activation with higher max-Z values, including regions in the ACC and bilateral frontal pole which are non-significant in the positively-valenced contrast. These results show that when immoral characters are punished, the between-subject mean functional connectivity patterns for rTPJ and PC are significantly different from when immoral characters are rewarded. But, on the other hand, when morally-upstanding characters are rewarded, the between-subject mean functional connectivity patterns are only marginally different from when moral characters are punished. This finding is theoretically consistent with the view outlined above (as well as in Weber et al., 2011) that people possess a particular innate sensitivity to the punishment of immoral characters. Given the results of the ISC analysis, it could be the case, for instance, that functional connectivity patterns are more individually-variant under positively-valenced conditions (yielding a null result when aggregated across all subjects), whereas negatively-valenced conditions brings these connectivity patterns into alignment across individuals by tapping into deeply-rooted intuitions (thereby producing many highly-significant results when aggregated across all subjects). Future research might be conducted to more clearly delineate the difference in these networks and generate more accurate predictions about how functional connectivity influences disposition preferences and narrative enjoyment.

In summary, the overarching purpose of this study was to examine the affective underpinnings of ADT using observational measures derived from brain activity and facial expressions. Consistent with prior research, ISC in brain activity is higher in disposition-consistent conditions than disposition-inconsistent conditions, and highest when immoral characters are punished. However, contrary to what was predicted, the affective component of ADT does not seem to be explicable on the basis of amygdala-vmPFC connectivity. The conceptual framework underlying Shenhav & Greene's (2014) work is that limbic and executive systems interact to produce an integrative moral judgment, and that higher connectivity between amygdala and vmPFC therefore represents the integration of affective response with reasoned judgment. Theory would suggest that stronger affective involvement should accentuate disposition-driven preferences; however, on an individual level, the co-activation of these particular seed regions did not predict the strength of affective disposition as measured by the ADT index (Weber et al., 2008). Future research may be well-served to consider the spatially-proximate and functionally-similar temporal pole, rather than the amygdala, given that PPI analysis revealed greater vmPFC-temporal pole connectivity (aggregated across all subjects) in the negatively-valenced contrast. Additionally, although the ISC results suggest the ability of negatively-valenced disposition-consistent content to align brain activity across individuals, the moderating role of trait empathy in the PPI analyses suggests that individual differences in empathy do affect functional connectivity patterns and should be accounted for in future models. The tension between these two findings is worthy of further study.

These results refute the prediction that correlations in facial expressions can serve as an indicator of ISCs in the brain. On the contrary, these results suggest that, if anything, higher facial correlations are associated with *lower* levels of ISC in relevant brain ROIs. While disposition-consistent content engages common patterns of brain activity across individuals, it is instead disposition-inconsistent content that engages common patterns of facial activity across individuals. As argued above, the most reasonable explanation for this effect seems to be that facial expressions do not serve as a direct window into a person's emotional state, but rather serve as a means to communicate salient information to others. Although it runs contrary to what was predicted, this finding leads to an intriguing direction for research into how this subtle form of nonverbal communication influences others watching the same narrative.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral Cognitive Neuroscience Review, 1*(1), 21—62.

Aramova, Y. R. & Inbar, Y. (2013). Emotion and moral judgment. *WIREs Cognitive Science, 4,* 169—178.

Barbey, A. K., Krueger, F., & Grafman, J. (2009). Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philosophical Transactions of the Royal Society B, 364,* 1291—1300.

Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). “I show how you feel”: Motor mimicry as a communicative act. *Journal of Personality and Social Psychology, 50*(2), 322—329.

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science, 5,* 187—202.

Berkman, E. T. & Falk, E. B. (2013) Beyond brain mapping: Using neural measures to predict real-world outcomes. *Current Directions in Psychological Science, 22*(1), 45 —50.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J.C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron, 60*(5), 930—940.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., & Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*, 205—2011. New York: ACM.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function, 217*(4), 783—796.

Damasio, A. (2000). A second chance for emotion. In R.D. Lane & L. Nadel (Eds.), *Cognitive Neuroscience of Emotion* (pp. 12—23). New York, NY: Oxford University Press.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.

Detert, J. R., Trevino, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology, 93*(2), 374—391.

Dinh, J. E. and Lord, R. G. (2013). Current trends in moral research: What we know and where to go from here. *Current Directions in Psychological Science 22*(5), 380—385.

Dufour, N., Redcay, E., Young, L., Mavros, P. L., Moran, J. M, Triantafyllou, C., … Saxe, R. (2013). Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLOSOne*, *8*(9), e75468.

Eckman, P. (1993). Facial expression and emotion. *American Psychologist, 48*, 384—392.ew

Fehr, E. & Gachter, S. (2002). Altruistic punishment in humans. *Nature, 415,* 137—140.

Firth, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions B, 364*(1535), 3453—3458.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping, 2,* 56—78.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396—403.

Gleichgerrcht, E. & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS One, 8*(4), e60418.

Goldman, A. I. and Sripada, C. S., (2005). Simulationist models of face-based emotion recognition. *Cognition, 94*(3), 193—213.

Green, J. D., Sommerville, R. B., Nystrom, L. E, Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105 —2108.

Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based recognition. *Neuroimage, 48*(1), 63-72.

Haidt, J. (2003). The emotional dog does learn new tricks: A reply to Pizarro and Bloom

(2003). *Psychological Review, 110,* 197—198.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., Keysers, C. (2012). Brain-to-brain coupling: mechanism for creating and sharing a social world. *Trends in Cognitive Science, 16*(2):114—121.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. Projections, 2(1), 1—26.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303, 1634—1640.

Kauppi, J.P., Pajula, J., & Tohka, J. (2014). A versatile software package for inter-subject correlation based analyses of fMRI. *Frontiers in Neuroinformatics, 8*, 2.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science, 342*(6), 377—380.

Lichev, V, Sacher, J., Ihme, K., Rosenberg, N., Quirin, M., Lepsein, J., … Suslow, T. (2015). Automatic emotion processing as a function of train emotion awareness: a fMRI study. *SCAN*, *10,* 680-689.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences, 11*(4), 143—152.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience, 6,* 799—809.

Moll, J. & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences, 11*(8), 319—321.

Narvaez, D. (2010). Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science 5*(2), 163—181.

Ochsner, K. N. & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9*, 242–249.

O'Donnell, M. B., & Falk, E. B. (2015). Linking neuroimaging with functional linguistic analysis to understand processes of successful communication. Communication Methods and Measures, *9*(1-2), 55—77.

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: A review of findings on social and emotional processing. *Brain, 130*, 1718-1731.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*(5), 604—609.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review, 9*(4), 278—311.

Pizarro, D. A. & Bloom, P. (2003). The intelligence of moral intuitions: Comment on Haidt (2001). *Psychological Review, 110*, 193—196.

Prinz, J. J. (2011). Is empathy necessary for morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 211—229). Oxford: Oxford University Press.

Raney, A. A. & Bryant, J. (2002). Moral judgment and crime drama: An integrated theory of enjoyment. *Journal of Communication*, *52*, 402—415.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review, 5*(1), 16—23.

Ross, D. (2007). H. sapiens as ecologically special: what does language contribute? *Language Sciences, 29*, 710—731.

Saragih, J. M., Lucey, S., & Cohn, J. (2009). Face alignment through subspace constrained mean-shifts. *IEEE International Conference on Computer Vision.* Retrieved from https://www.ri.cmu.edu/pub_files/2009/9/CameraReady-6.pdf.

Schaich-Borg, J., Lieberman, D., and Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience, 20*, 1529—1546.

Shenhav, A. & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience, 34,* 4741—4749.

Singer, T., Seymour, B., O'Doherty, J. P., Klass, S. E., Dolan, R. J., & Firth, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466—469.

Stephens, G. J., Silbert, L. J., Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceeding National Academy of Science USA, 107*(32) 14425—14430.

Tamborini, R. (2011). Moral intuition and media entertainment. *Journal of Media Psychology: Theories, Methods, and Applications 23*(1), 39—45.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59,* 204—217.

Vogt, T., André, E., & Bee, N. (2008). EmoVoice: A framework for online recognition of emotions from voice. *Lecture Notes in Computer Science, 5078*, 188—199.

Weber, R., Eden, A., & Mathiak K. (2011). *Seeing bad people punished makes us think alike: Social norm violations in television drama elicit cortical synchronization in viewers.* Paper presented at the annual meeting of the International Communication Association, Boston, MA.

Weber, R., Sherry, J., & Mathiak, K. (2008). The neurophysiological perspective in mass communication research: Theoretical rationale, methods, and applications. In M. J. Beatty, J. C. McCroskey, & K. Floyd (Eds.), *Biological Dimensions of Communication: Perspectives, Methods, and Research* (pp. 41—71). Cresskill, NJ: Hampton Press.

Weber, R., Tamborini, R., Lee, H. E., & Stipp, H. (2008). Soap opera exposure and enjoyment: A longitudinal test of disposition theory. *Media Psychology*, *11*, 462—487.

Zillmann. D. (2006). Dramaturgy for emotions from fictional narration. In J. Bryant & P. Vorderer, *Psychology of Entertainment*. Mahwah, NJ: Erlbaum.

Zillmann, D. (2013). Moral monitoring and emotionality in responding to fiction, sports, and the news. In R. Tamborini, (Ed.), *Media and the Moral Mind*. New York: Routledge.

Zillmann, D. & Cantor, J. R., (1977). Affective responses to the emotions of a protagonist. *Journal of Experimental Social Psychology, 13*(2), 155—165.

APPENDIX A

FACE-TRACKING SOURCE CODE

Main script, process_video.py:

```
#!/usr/bin/env python
import json,os,subprocess,sys,time,threading,subprocess
from collections import OrderedDict,deque
from partool import cluster_list,queue_list
from constants import *
from CGIHTTPServer import CGIHTTPRequestHandler
from BaseHTTPServer import HTTPServer

CHROMECALL = 'chromium http://localhost:9999/{i} &'
TEMPLATE = VIDEOPATH+'process_template.html'

# run http server as thread
os.chdir(VIDEOPATH)
adr = ('',9999)
httpd = HTTPServer(adr, CGIHTTPRequestHandler)
threading.Thread(target=httpd.serve_forever).start()

with open(JSONPATH+'order_data.json','r') as jsonfile:
    order_data =
json.load(jsonfile,object_pairs_hook=OrderedDict)

with open(TEMPLATE) as templatefile:
    template = templatefile.read()

cmdlist = deque()
for subj in order_data.keys():
    for video in range(1,16):
        vid = str(video).zfill(2)
        uid = subj.zfill(2)
        # make the html file
        of = template.format(s=uid,v=vid)
        ofile = '{}html/{}_{}.html'.format(VIDEOPATH,uid,vid)
        with open(ofile,'w') as outfile:
            outfile.writelines(of)
        # run the html file
        i = 'html/{}_{}.html'.format(uid,vid)
        cmd = CHROMECALL.format(i=i)
        cmdlist.append(cmd)

while len(cmdlist) > 0:
    subprocess.call(cmdlist.popleft(),shell=True)
    # wait awhile so they don't pile up
    time.sleep(195)
    # kill it
    subprocess.call('pkill chromium',shell=True)
    time.sleep(5)
httpd.shutdown()
```

Contents of process_template.html:

```
<html>
  <head>
    <script src="http://localhost:9999/clmtrackr.js"></script>
    <script
src="http://localhost:9999/model_pca_20_svm_emotionDetection.j
s"></script>
    <script
src="http://localhost:9999/emotionmodel.js"></script>
    <script
src="http://localhost:9999/emotion_classifier.js"></script>
  </head>
  <body>
    <div>
    <video id="inputVideo" width="720" height="480">
      <source src="http://localhost:9999/thinkaloud/trimmed/
{s}_{v}.ogv" type="video/ogg"/>
    </video>
    <canvas id="drawCanvas" width="720" height="480"
position="float"></canvas>
    </div>
    <div id="data1"></div>
    <div id="data2"></div>
    <div id="data3"></div>
    <script
src="http://localhost:9999/process_video.js"></script>
    <form action="http://localhost:9999/cgi-bin/write.py"
method="POST" id="subform">
      <input type="text" name="video" id="video"
value="{s}_{v}"/>
      <textarea name="jsondata" id="jsondata"></textarea>
    </form>
  </body>
</html>
```

Contents of process_video.js:

```
var input = document.getElementById('inputVideo');
var data1 = document.getElementById('data1');
var data2 = document.getElementById('data2');
var data3 = document.getElementById('data3');
function finished(e) {
    clearTimeout(dataloop);
    ctracker.stop();
    outfield = document.getElementById('jsondata');
    outfield.value = '['+outdata.toString()+']';
    document.getElementById('subform').submit();
}
var outdata = [];
function log_data() {
    position = ctracker.getCurrentPosition();
    params = ctracker.getCurrentParameters();
    emotion = ec.meanPredict(params);
    data1.innerHTML = position;
    data2.innerHTML = params;
    if (emotion) {
        emotion_string =
emotion[0].emotion+emotion[0].value.toString()
+emotion[1].emotion+emotion[1].value.toString()
+emotion[2].emotion+emotion[2].value.toString()
+emotion[3].emotion+emotion[3].value.toString();
        data3.innerHTML = emotion_string;
    }

outdata.push(JSON.stringify({"position":position,"parameters":
params,"emotions":emotion}));
}
input.addEventListener('ended',finished,false);
var ctracker = new clm.tracker();
ctracker.init(pModel);
ec = new emotionClassifier();
ec.init(emotionModel);
input.play();
ctracker.start(input);
dataloop = setInterval(log_data,500);
var canvasInput = document.getElementById("drawCanvas");
var cc = canvasInput.getContext('2d');
function drawLoop() {
    requestAnimationFrame(drawLoop);
    cc.clearRect(0,0,canvasInput.width,canvasInput.height);
    ctracker.draw(canvasInput);
}
drawLoop();
```

Contents of write.py (CGI script to save output):

```
#!/usr/bin/env python
import cgi

FACEPATH = '/mnt/diss/predict/face/'

form = cgi.FieldStorage()
vid = form.getvalue('video')
filename = '{}{}'.format(FACEPATH,vid)

with open(filename,'w') as outfile:
    outfile.write('{}\n'.format(form.getvalue('jsondata')))

print 'Content-type:text/html\n\n'
print '<html><body>Wrote data to
<strong>{}</strong></body></html>'.format(filename)
```

Video pre-processing script trim_video.py:

```
#!/usr/bin/env python

from collections import defaultdict, OrderedDict
from partool import queue_list

INVIDEODIR =
'/home/jmm/diss_videos/thinkaloud/InterviewVideos/'
OUTVIDEODIR = '/home/jmm/diss_videos/thinkaloud/trimmed/'
MENCODERCALL = 'mencoder -ss 00:00:{start} -endpos 00:03:00
-oac copy -ovc copy {idir}{subj}_{video}_original.mpg -o
{odir}{subj}_{video}.mpg'
MAPFILE = '/home/jmm/diss_videos/video_timing.txt'

vdata = defaultdict(OrderedDict)

with open(MAPFILE,'r') as mapfile:
    for ln in mapfile:
        arr = ln.rstrip().split('\t')
        if len(arr) < 3:
            continue
        subj = arr[0]
        video = arr[1]
        start = arr[2]
        if start.endswith('*'):
            start = start[0]
        vdata[subj][video] = start

cmdlist = []
for subj,data in vdata.items():
    for video,start in data.items():
        s = subj.zfill(2)
        v = video.zfill(2)
        t = start.zfill(2)
        cmd =
MENCODERCALL.format(idir=INVIDEODIR,odir=OUTVIDEODIR,start=t,s
ubj=s,video=v)
        cmdlist.append(cmd)

queue_list(cmdlist,threadcount=4)
```

Video conversion script convert_video.py:

```
#!/usr/bin/env python

import json
from constants import *
from partool import queue_list
from collections import OrderedDict

CONVERTER = 'ffmpeg -i {s}_{v}.mpg {s}_{v}.ogv'

with open(JSONPATH+'order_data.json','r') as jsonfile:
    order_data =
json.load(jsonfile,object_pairs_hook=OrderedDict)

cmdlist = []
for subj in order_data.keys():
    for i in range(1,16):
        subz = subj.zfill(2)
        iz = str(i).zfill(2)
        cmd = CONVERTER.format(s=subz,v=iz)
        cmdlist.append(cmd)
```

```
queue_list(cmdlist,threadcount=4)
```