

Towards a neuroscience of compassion: A brain systems-based model and research agenda

Yoni K. Ashar¹

Jessica R. Andrews-Hanna²

Sona Dimidjian¹

Tor D. Wager^{1,2}

¹Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder,
CO

²Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO

Please address correspondence to:

Yoni K. Ashar

Department of Psychology and Neuroscience

University of Colorado, Boulder

345 UCB

Boulder, CO 80309

Email: yoniashtar@gmail.com

Abstract

Despite substantial progress in the last decade towards understanding the neural building blocks of empathy, relatively little is known about the neural bases of *compassion* – a complex internal state characterized by prosocial motivation to improve the other’s condition. In Parts I and II of this chapter, we integrate existing literature on empathy, altruism, and social cognition to develop a neuropsychological process-content model of compassion and compassionate behavior. In this model, compassion is comprised of multiple component processes, including the generation of affective feelings, inferences about others’ mental states, and appraisal of the meaning of another’s suffering in relation to oneself. These component processes are supported by distinct brain systems, which represent content—specific feelings, judgments, and meaning representations—in the form of unique spatio-temporal patterns of neural activity. Like an “attractor network,” these activity patterns dynamically interact both within and across networks, leading to system-wide configurations of network activity that characterize the response to the suffering individual.

In Part III, we use our dynamic process-content model of compassion as a framework for suggesting important future directions for compassion research. We highlight the promise of compassion training interventions to enhance prosocial behavior, and we call for translational research leveraging the tools of cognitive neuroscience to illuminate the mechanisms of compassion training. We conclude by applying our model of compassion to some of our own recent research.

Keywords: compassion, altruism, empathy, compassion training, social cognition, emotion, meditation

“Love and compassion are necessities, not luxuries. Without them humanity cannot survive.”

– *The Dalai Lama*

Compassion is regarded as a central virtue by many cultures and value systems. It is an essential ingredient of healthy interactions with others at every scale, from the everyday interactions within a local community to the interactions among nations that shape human wellbeing and suffering in profound ways. Though compassion is interpersonal, it has also been empirically linked with personal benefits, including increased positive emotions (Dunn, Aknin, & Norton, 2008), improved physical health (Carson et al., 2005; Kok et al., 2013), and a reduced immunological stress response (Pace et al., 2010). Unfortunately, despite the personal and interpersonal advantages of a compassionate stance, people often respond to others' suffering with indifference, aversion, or even gloating. A scientific understanding of *how* and *when* compassion arises could help promote a more compassionate society.

To illustrate the complexity of responding to other's suffering, imagine encountering a disheveled elderly woman begging for money on the street corner. Perhaps she seems desperately needy and frantic, or perhaps she seems jaded and worn from years of begging. Basic affective feelings—the desire to approach or avoid, elementary forms of distress, tenderness, and aversion—arise immediately, often unbidden. Simultaneously, an assortment of thoughts may present themselves, such as: It's her fault she's homeless, she probably has nowhere to sleep, my \$2 won't do her much good anyways, she seems like a sweet person, I can't trust her to spend the money wisely, and so forth. In some cases, this information is integrated to construct a schematized, gestalt “emotional meaning”¹ (Roy, Shohamy, & Wager, 2012) regarding the situation, such as compassionately perceiving the woman as deserving of

help or angrily blaming her for her suffering. All of these feelings, judgments, and emotional meanings interact, and each can constrain the evolution of the others, in a process potentially resulting in behavioral decisions such as helping or distancing. Understanding these interactions, particularly the factors that lead to the evolution of compassion vs. disgust or schadenfreude, could support the development of targeted interventions to increase compassion and compassionate behavior.

There are many obstacles to studying compassion, not least among them that scholars define compassion in different ways, and the boundaries between affective feelings, judgments, and emotions are conceptually permeable. The concept of “compassion” can potentially include feelings, thoughts, emotions, and behaviors (helping). Here, we operationally define compassion as the motivation to relieve the suffering of another. The conceptual ambiguity inherent in defining psychological processes such as “compassion” and “emotion” is a major reason to anchor concepts and definitions in the study of brain systems. Mapping compassion and its psychological ingredients to brain systems can provide a stable framework for identifying processes independent of semantic definitions, and a basis for their objective measurement.

In Part I of this chapter, we emphasize that compassion and compassionate behavior comprise multiple component processes, including affective feelings, social inferences, and emotional meanings, each supported by distinct brain networks. In Part II, we describe the relationships between these component processes, illustrating how the attractor-like properties of their underlying brain networks facilitate a dynamic interplay of patterns of neural activity. In Part III of this chapter, we identify several future directions for the field of compassion research, focusing especially on compassion-training interventions. We conclude with recent findings

from our own research that seeks to clarify the basic neuropsychological underpinnings of compassion, compassionate behavior, and compassion training.

<1> Part I: Neural underpinnings of compassion

A growing body of research suggests that at least two distinct neural networks underlie empathy, the sharing and understanding of another's experience (de Waal, 2008; Decety & Jackson, 2004; Fan, Duncan, de Greck, & Northoff, 2011; Lamm, Decety, & Singer, 2010; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009; Van Overwalle & Baetens, 2009; Zaki & Ochsner, 2012). One network comprising the dorsal medial prefrontal cortex (dmPFC), temporoparietal junction (TPJ), and posterior cingulate cortex (PCC) supports *social-inferential* properties of empathy, as when inferring the perspectives, beliefs, and feelings of other people. A second, distinct network centered on the anterior insula (aI) and the dorsal anterior cingulate cortex (dACC) engages when individuals experience *affective* responses to others' suffering.

While these brain systems support sharing in and understanding another's suffering, a distinct brain system underlies the valuing of others and prosocial motivation to help them (Goetz, Keltner, & Simon-Thomas, 2010; Singer & Lamm, 2009). We posit that compassion is supported by a medial prefrontal-striatal network constructing (potentially compassionate) emotional meanings (Roy et al., 2012).

<2> Social Inference. Compassion and compassionate behavior depend first and foremost on perceiving the other to be in need (Batson, 2011). Additionally, compassion and compassionate behavior have been empirically shown to depend on a number of other social inferences, including attributions of responsibility (Rudolph, Roesch, Greitemeyer, & Weiner, 2004), trustworthiness (Sargeant & Lee, 2004; van't Wout & Sanfey, 2008), likability (Batson &

Lishner, 2005), and in some cases self-similarity (Batson & Lishner, 2005; Vollhardt & Staub, 2011). Social inferences depend on the ability to understand and make attributions regarding others' mental states, a process often referred to as "mentalizing" or "theory of mind". A system of cortical structures including the dorsal medial prefrontal cortex (dmPFC), posterior cingulate cortex (PCC), and temporoparietal junction (TPJ) (*Fig. 1A*, blue network) is widely thought to support these processes (Frith & Frith, 2006; Lieberman, 2007; Mitchell, Macrae, & Banaji, 2006). This network is recruited in explicitly compassion-relevant processes, such as rating the intensity of others' emotional pain (Bruneau, Dufour, & Saxe, 2012; Bruneau, Pluta, & Saxe, 2012), passing moral judgment on others (Koster-Hale, Saxe, Dungan, & Young, 2013), and accurately inferring others' emotions (Zaki, Weber, Bolger, & Ochsner, 2009).

<2> Affective Feeling. A second system, including the ventral mid-anterior insular cortex (aI), dorsal anterior cingulate (dACC), and their connections with the amygdala, supports affective responses to others' suffering (*Fig. 1A*, red network) (Barrett & Satpute, 2013; Chang, Yarkoni, Khaw, & Sanfey, 2012; Duerden, Arsalidou, Lee, & Taylor, 2013; Zaki & Ochsner, 2012). Affective feelings here refer to basic, rudimentary feelings, with motivational properties, but without elaborated conceptual schemas (Russell & Barrett, 1999).

Two types of affective responses, related but conceptually distinct, may arise in response to others' suffering: a) affective responses directly to the stimulus itself, which are perhaps evolutionarily prepared, such as feeling distress when hearing a child cry (termed "distress for" by Batson, 2011), and b) the vicarious experience of another's internal state ("distress as" or "distress with"). The mechanism for this latter type of shared affective response is hypothesized to be a mirror neuron network, in which the same networks engaged during first-hand experience of affect also subserve empathic responses (de Waal, 2008; Decety & Jackson, 2004; Engen &

Singer, 2012; Gallese & Goldman, 1998; Singer & Lamm, 2009). Further research is needed to disambiguate these two affective responses to others' suffering, especially since they may have different motivational consequences regarding helping behavior.

The relationship between affective feeling and compassionate behavior is complex. The interpersonal implications of an emotion, rather than its basic affective properties, seem to be most predictive of compassion. For example, tenderness, sadness, and concern have been linked with prosocial motivation and behavior (Batson, Fultz, & Schoenrade, 1987; Eisenberg, Fabes, & Miller, 1989) as well as personal distress (unpublished data, described below), suggesting that valence and arousal alone are not linearly related to helping. Relatedly and somewhat paradoxically, Condon and Barrett found that compassion is conceptualized as a positively valenced emotion, but the experience of compassion leads to heightened negative affect (Condon & Barrett, 2013). Further research is needed to unpack the motivational consequences of various affective responses, focusing especially on the interpersonal consequences of emotions.

<2> Emotional Meaning. Compassion is often characterized by a schematized emotional appraisal of the suffering other, informed by the suffering other's personal significance to the self. We propose that a ventromedial prefrontal-subcortical network (*Fig. 1A*, green network) subserves this process, which we describe as the construction of "emotional meaning" (Roy et al., 2012). The ventromedial prefrontal cortex (vmPFC) connects systems involved in episodic memory, representation of the affective qualities of sensory events, social cognition, and interoceptive signals, and plays a unique role in representing conceptual information and in transducing concepts into affective behavioral and physiological responses (Haber & Knutson, 2010). Additionally, the vmPFC has the requisite connections to the social inference and associative affect networks outlined above, as well as to the striatum, hypothalamus, and

brainstem, allowing it coordinate system-wide affective physiological and behavioral responses (Roy et al., 2012). This network shows increased activity when participants are asked explicitly to adopt a compassionate stance toward others' suffering (Kim et al., 2009; O. M. Klimecki, Leiberg, Lamm, & Singer, 2012), and connectivity within this network correlates with sadness when viewing a film about another's suffering (Raz et al., 2012).

The emotional meaning constructed around another's suffering includes how much a person "cares" about the suffering other, reflecting psychological processes of valuing the other (Batson, 2011) and evaluating the other's relevance for the self (Goetz et al., 2010).

Neuroimaging studies confirm that activity in vmPFC-subcortical circuits track the closeness of another person (Krienen, Tu, & Buckner, 2010), even when closeness is crossed with valence: for Arabs and Jewish Israelis, considering the suffering of in-group or out-group members activated the vmPFC equally, while this region showed relatively less activity when considering South Americans (a distant group) (Bruneau, Dufour, et al., 2012).

Importantly, helping behavior will often reflect the emotional meaning constructed. Neuroimaging has revealed that activity in the vmPFC and ventral striatum is associated with the size of charitable donations participants make (Harbaugh, Mayr, & Burghart, 2007; Hare, Camerer, Knoepfle, & Rangel, 2010; Moll et al., 2006), decisions to give equitably (Zaki & Mitchell, 2011), and prosocial behavior towards a socially-excluded other (Masten, Morelli, & Eisenberger, 2011).

<2> Additional brain systems. Though not the focus of the present review, a number of additional brain systems may also play an important role in generating a compassionate response to a suffering individual. If the individual is a familiar other, neural circuits subserving memory retrieval, including the medial temporal lobe, retrosplenial cortex, and

posterior inferior parietal lobule, may facilitate the retrieval of prior information relevant to the individual. Fronto-parietal control systems, including those involved in emotion regulation, may additionally help the observer inhibit his pre-potent emotional or behavioral response, or resolve his internal conflict regarding the possible courses of action. Additionally, the “mirror neuron” system, including the intraparietal sulcus, posterior superior temporal sulcus (pSTS), and premotor cortex, plays an important role in understanding others’ motor goals and actions (Rizzolatti & Sinigaglia, 2010; Van Overwalle & Baetens, 2009) and is important for several forms of empathy (Fan et al., 2011; Lamm et al., 2010).

<1> Part II: A dynamic process-content model of compassion

Thus far, we have described three distinct brain networks, each implicated by prior research as supporting a category of processes critical for compassion to arise. We now describe the interrelationship of these processes, focusing especially on the *content* of these processes, a feature sometimes overlooked in neuroimaging research.

In the model, a target of compassionate behavior (i.e. a suffering woman) can be processed both in a feed-forward fashion, in which information flow through the brain is primarily unidirectional (*Fig. 1B*, left-to-right directionality) and in an iterative, recurrent fashion, like an attractor network (*Fig. 1B*, within- and between-network arrows), where information flows multi-directionally within and between networks.

After early visual processing of the stimulus (a suffering individual), one or many affective feelings may arise. The precise nature of these feelings (i.e. their *content*) may be represented as unique *patterns* of neural activity within the insular-cingulate network. Simultaneously, the observer may form social inferences about the mental state and condition of

the suffering individual, represented by distinct patterns of neural activity within the network comprising the dmPFC, TPJ, and PCC. The patterns of activity in these two networks are next integrated in the medial prefrontal-striatal network, along with additional information, to form an emotional meaning of the suffering individual. The strength and nature of the emotional meaning will ultimately guide the observer's behavior.

[INSERT FIGURE 1 HERE]

The networks supporting social inference and affective feelings act as “rate-limiting processes” (in the language of biology), such that diminished function in either will greatly reduce the likelihood of a compassionate emotional meaning arising. Likewise, helping behavior toward the other is unlikely unless a critical intensity of emotional meaning is reached. Thus, the presence of these three processes—feelings, social inference, and meaning—is necessary for compassion and compassionate behavior. The absence of these processes would be indicative of apathy.

Each network functions as an attractor network: the activity of any one population of neurons impacts other populations of neurons, causing changes to iteratively reverberate throughout the network, such that the network will progress through different states. At a system-wide level, the three networks also function as an attractor network: activity in one network will impact the other two, which will then impact the first network, etc., evolving until a stable system-wide pattern potentially emerges. Thus, quicker responses (such as a startle response or fast aversion) and their concordant neural activations may yield to slower, potentially trained responses, including compassion.

To illustrate the proposed model, we return to our opening example. Upon first sight of the homeless woman, feelings of aversion and/or a desire to engage are represented as fast,

competing patterns in an amygdala/insula/ACC network. In tandem, social inferences—such as she is suffering, she wants me to give her money, she is hungry, etc.—arise as different patterns in the dmPFC/PCC/TPJ network. Within each network, populations of neurons progress through between different states, representing these competing or coexisting social inferences and affective feelings. An emotional meaning emerges as the individual constructs a situated, gestalt representation of the situation congruent with the woman's personal significance to the individual, such as 'this poor woman needs my help' or 'this woman is disgusting to me and deserves her suffering'. The content in different networks may mutually influence each other. For example, feelings of disgust may strengthen social inferences of blame, while a simultaneous prosocial emotional meaning will strengthen competing social inferences of blamelessness, and activity in all three networks will adapt accordingly. Compassion and other distinctive responses such as gloating are characterized by the formation of a stable, coherent, system-wide representation of the feelings, thoughts and narrative surrounding the encounter, leading to coordinated behaviors including helping. Conversely, an individual may continue to feel torn, characterized by continuous oscillations between various network configurations, and directed behavior may not emerge.

<2> Implications for compassion deficits

Our model of compassion is consistent with established findings regarding the dissociable neural bases of compassion deficits in specific clinical populations. Psychopathy, schizophrenia, depersonalization and narcissism are characterized by deficits in affective feeling but not necessarily social inference, while autism, bipolar disorder and borderline traits are associated with impairment in social inference but not affective feeling (reviewed in Cox et al., 2012).

Similar mechanisms may underlie the dehumanization of devalued others. In dehumanization, compassion may fail to arise because of difficulty inferring others' internal states (Harris & Fiske, 2007; Haslam, 2006), because affective responses are suppressed, or because the content of affective responses does not support compassion. Dehumanization can be supported by distorted emotional meanings as well, such as narratives depicting others as sexual objects (Bernard, Gervais, Allen, Campomizzi, & Klein, 2012) or as "parasites" and existential threats (Herf, 2006).

<1> Part III: Compassion training and other future directions

Many open, important questions remain in compassion research. Here we highlight a few prominent future directions, with an eye toward compassion training (CT) due to its potential for large-scale use and its clear societal implications.

<2> Linking cognitive neuroscience with compassion training. Recent evidence suggests that people can be trained both to feel and act more compassionately towards others. Compassion training (CT) programs have typically been based on Compassion Meditation and/or Loving-Kindness Meditation, contemplative practices in which the practitioner practices feeling care, connection, and love for others and/or reflects on others' suffering and human interdependence (Salzberg, 2002). Relative to a variety of control conditions, CT has been shown to increase prosocial behavior in a video game (Leiberg, Klimecki, & Singer, 2011), accuracy in discerning others' emotions (Mascaro & Rilling, 2012), self-reported empathy (O. M. Klimecki et al., 2012), altruistic redistribution of funds to benefit others treated unfairly (Weng et al., 2013), and real-world helping behavior to strangers in need (Condon, Desbordes, & Miller, 2013). Evidence is also accumulating that CT leads to changes in neural function, including

enhanced neural activity in the right amygdala (Desbordes et al., 2012), inferior parietal lobule and dorsolateral PFC (Weng et al., 2013), and striatum and right medial orbitofrontal cortex (O. M. Klimecki et al., 2012) when witnessing human suffering. CT has also been shown to increase neural activity in the inferior frontal gyrus and dorsomedial PFC when attempting to infer others' emotions (Mascaro & Rilling, 2012). Relatedly, expert compassion meditators showed enhanced neural processing in the TPJ, posterior superior temporal sulcus, amygdalae, and insula in response to sounds of human distress (Lutz, Brefczynski-Lewis, Johnstone, & Davidson, 2008). The diversity of brain regions that have shown a response to CT likely reflects the underlying diversity of compassion training and measurement paradigms employed.

Thus, while the evidence indicates that compassionate behavior and related neural function can be trained, the psychological and neural mechanisms mediating this change remain unclear. To our knowledge, studies have not yet identified specific psychological processes mediating the behavioral changes induced by CT, and the marked heterogeneity of brain changes resulting from CT indicates ambiguity regarding its neural mechanisms. Advances in the basic science of compassion will pave the way for CT interventions to more precisely target and assess specific processes and neural networks, enabling robust, replicable interventions.

Using our dynamic model of compassion as a platform, we suggest that CT may increase prosocial behavior by targeting the component processes highlighted above, in addition to the underlying nature of their content. For example, a CT intervention might aim to enhance prosocial inferences, positive affective feelings and/or empathetic suffering, and more compassionate emotional meanings. However, our model also predicts that an intervention targeting only one system is likely to affect the other systems as well, given the interactive nature of these systems.

Our model also has implications for how compassion-related brain activity can be detected and measured using Functional Magnetic Resonance Imaging (fMRI) and other technologies. While overall regional activity may distinguish between compassion and apathy (i.e., Harbaugh et al., 2007; Hare et al., 2010; Klimecki et al., 2012), distinguishing between different responses of similar intensities (such as compassion and schadenfreude) will require examining within-network patterns of activity, rather than overall regional activity. For example, evidence is accumulating that both compassion (Harbaugh et al., 2007; O. M. Klimecki et al., 2012) and schadenfreude (Dvash, Gilam, Ben-Ze'ev, Hendlar, & Shamay-Tsoory, 2010; Hein, Silani, Preuschhoff, Batson, & Singer, 2010) activate the striatum, a component of the emotional meaning network. Likewise, Arabs and Jewish Israelis report feeling less compassion for the suffering of an out-group member, but showed no differences in overall regional neural activity in the vmPFC, PCC, or rTPJ compared to the in-group member (Bruneau, Dufour, et al., 2012), suggesting that within-region *patterns* of neural activity may be more informative than overall activity levels in some cases. Consequently, multi-voxel pattern analysis (MVPA), a multivariate technique allowing detection of spatially distributed patterns of activation, may be a useful tool in differentiating between different, but equally intense, responses to others' suffering.

Additionally, our model posits a theoretical mechanism by which training occurs: As the networks practice stabilizing in compassionate states, these prosocial patterns will become more readily accessible and adopted more quickly. This prediction could be explored by measuring the time delay for a participant to reach a "compassionate" neural state, potentially quantified using MVPA analyses that identify spatial patterns of activation linked to psychological processes.

<2> Improved measurement of compassion and compassionate behavior. Investigations of compassion and compassionate behavior are limited by the validity of the tools we use to measure those phenomena. A common approach is collecting self-reported compassion, but this approach is vulnerable to demand characteristics both from the specific experimental context and from the general social desirability of adopting a compassionate stance (DellaVigna, List, & Malmendier, 2012; Izuma, Saito, & Sadato, 2010). Facial expression may be valuable for measuring compassion (Eisenberg et al., 1989), particularly with the recent advent of automated facial expression recognition software (i.e. Littlewort et al., 2011). In addition, neuroimaging results can be most confidently interpreted in the context of in-scanner behavior, such as in-scanner charitable donation (Harbaugh et al., 2007; Hare et al., 2010) or empathic accuracy tasks (Mascaro & Rilling, 2012; Shamay-Tsoory & Aharon-Peretz, 2007). Lastly, real-world measures of compassionate behavior in which the participant is unaware of being observed will serve an important role in improving the ecological validity of compassion assessment. Some examples of ecologically-valid paradigms include whether a person gives up his or her seat to another person on crutches (Condon et al., 2013) and audio sampling of daily life to assess compassionate speech and behavior (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001).

<2> Translating compassion into compassionate behavior. Compassion will benefit others to the extent that it leads to compassionate behavior. Individuals feeling compassion will choose not to help if the costs associated with helping outweigh the anticipated benefits (Batson, 2011). Since the material costs of helping are small in many situations (i.e., \$2, five minutes of one's time, etc.), it is likely that other social, cognitive, or emotional costs likely deter helping behavior. For example, the presence of other individuals ("bystander effect") has been robustly demonstrated to hinder helping behavior (reviewed in Fischer et al., 2011). Relatedly, social

pressure can increase helping (DellaVigna et al., 2012; Izuma et al., 2010). Further research examining factors facilitating and inhibiting the behavioral expression of compassion, and the responsiveness of those factors to training, will be of great value to intervention research.

<2> Sustaining compassion. The “burning out” of individuals in care-giving positions is a well-documented, costly phenomenon (Coetzee & Klopper, 2010; Najjar, Davis, Beck-Coon, & Carney Doebbeling, 2009). It is possible that certain ways of empathizing will lead to burn out, while others will not (O. Klimecki & Singer, 2012). For example, personal distress over another’s suffering may exhaust a caregiver’s emotional resources, while feelings of tenderness and love might not, as some evidence suggests that personal distress motivates escape rather than helping behavior (Batson et al., 1987; Eisenberg et al., 1989). Potentially, interventions could train caregivers to adopt sustainable ways of empathizing. This is a ripe area of research with potentially direct applications.

<2> Our work. Initial results from our research show promise for linking the basic science of compassion with intervention research and identifying the mechanisms of CT. In a series of studies employing behavioral, neuroimaging, and intervention-based approaches, we identified several social inferences and emotional meanings that successfully predicted compassionate behavior. We operationalized compassionate behavior as charitable donation, where participants could choose to donate their own experimental earnings to real people in need. In an initial behavioral study, we used exploratory factor analyses to partition a wide variety of affective feelings, social inferences, and emotional meanings into distinct content dimensions reflecting tenderness, personal distress, perceived neediness, blaming, likability and perceived self-similarity. Collectively, these dimensions explained much of what motivated participants’ donation decisions, accounting for 41% of the variance in trial-by-trial donation amounts, within

person (unpublished data). Each of these content dimensions significantly and independently predicted donation amounts, except for self-similarity, which was not significant when controlling for other factors. In a second study, we compared a four-week CT intervention against two control conditions. We found that training-induced changes in tenderness, personal distress, perceived neediness, perceived blame, and likeability significantly mediated changes in compassionate behavior (unpublished data), suggesting these factors may be psychological mechanisms by which CT influences behavior.

In an additional neuroimaging study, we explored the neural underpinnings of these content dimensions. Initial findings show that tenderness and personal distress both correlate with increased activity in the ventromedial prefrontal cortex (vmPFC) (unpublished data), consistent with this region's role in constructing emotional meaning. This finding supports the notion proposed above that different emotional meanings are encoded in the vmPFC and will be best distinguished by within-region patterns of activation rather than overall regional activity, as remains to be investigated in future analyses. Further, we found that the amount of charitable donations offered by participants correlated with an overlapping region of the vmPFC, consistent with the proposed link between emotional meaning and prosocial behavior. We believe this work represents an important contribution to the development of a quantitative, biologically informed model of compassion and compassion training.

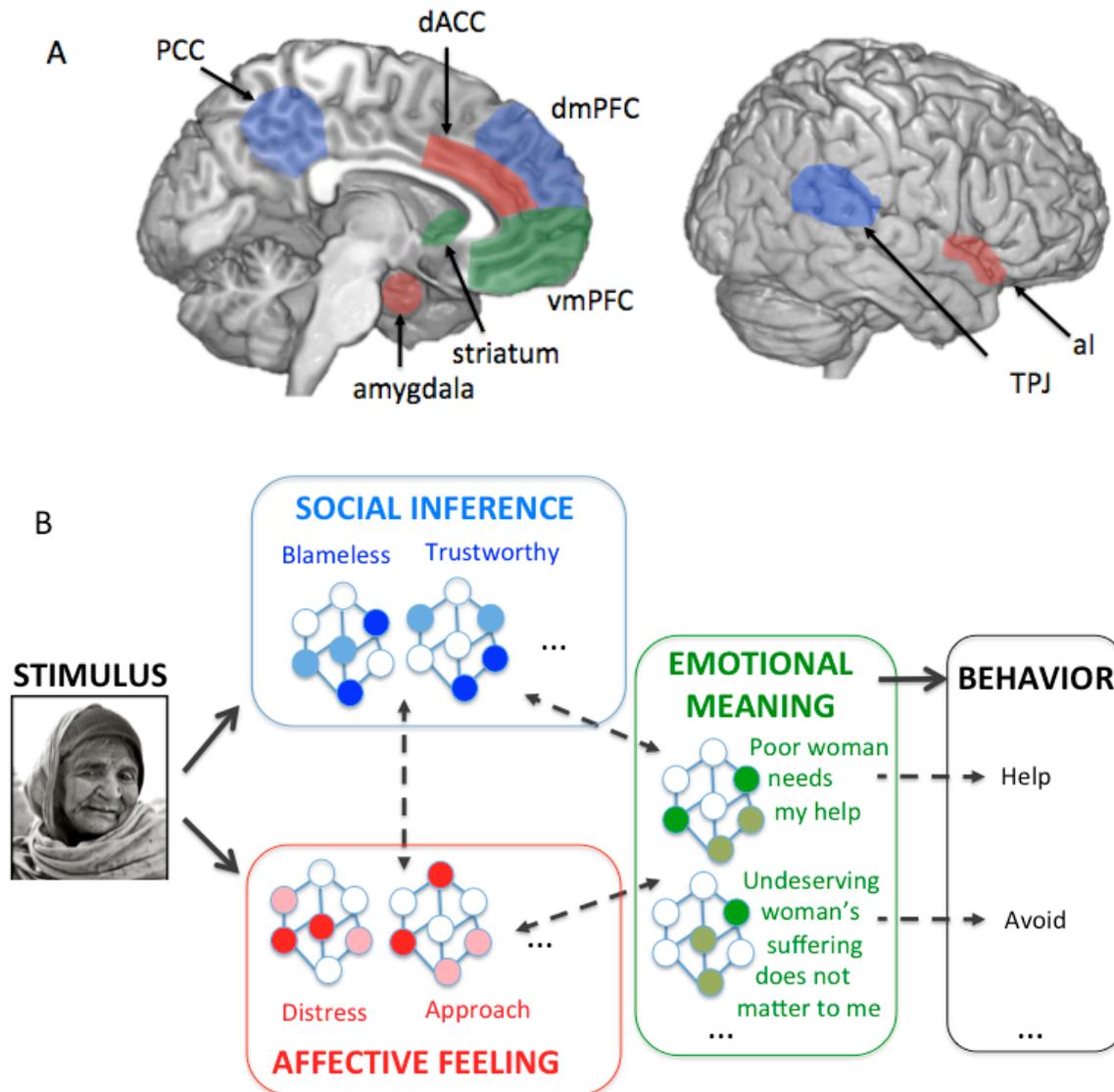
<1> Conclusions

We have proposed here a dynamic, process-content model of compassion, whereby social inference and affective feelings support the construction of potentially compassionate emotional meanings regarding a suffering other. This model also functions as a dynamic system, like an

attractor state network, where patterns of neural activity both within and across networks engage in a dynamic, interactive process, potentially leading to behavioral decisions. Importantly, advancing the basic science of compassion, as we have tried to do here, can empower intervention research. Although accumulating evidence suggests that compassionate feeling, behavior, and neural function do respond to training, the neuropsychological mechanisms supporting these changes are unclear. By carefully targeting and measuring specific neuropsychological processes, we will ultimately be able to design more powerful, robust, and generalizable CT interventions, building toward a kinder world.

Figure 1A. Key brain systems hypothesized to support compassion. One network of regions (red) supports *affective feelings* such as aversion or approach motivation, resulting either from sharing the affective state of the other or reacting directly to the stimulus. A distinct network (blue) supports *social inferences* and attributions regarding the other's internal state, such as 'she is poor' or 'it's her fault she is poor'. A third network (green) synthesizes conceptual and affective information to evaluate the gestalt, situated *emotional meaning* of the stimulus (Roy et al., 2012). aI, anterior insula; TPJ, temporoparietal junction; vmPFC, ventromedial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; PCC, posterior cingulate cortex; dACC, dorsal anterior cingulate cortex. Regional boundaries are approximate. **Figure 1B.** The suffering of another person is initially represented as social inferences and affective feelings. Information in these two networks contributes to the formation of an emotional meaning, which may guide the observer's behavior. These systems can be conceptualized as attractor networks, such that dynamic, interacting patterns of activity both within and between networks characterize the possible responses to other's suffering. With training, these networks can learn to adopt prosocial patterns more readily.

Figure 1:



<1> References

- Barrett, L., & Satpute, A. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current opinion in neurobiology*.
- Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.
- Batson, C. D., Fultz, J., & Schoenrade, P. a. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of personality*, 55(1), 19–39.
- Batson, C. D., & Lishner, D. A. (2005). Similarity and Nurturance: Two Possible Sources of Empathy for Strangers. *Basic and Applied Social Psychology*, 27(1), 15–25. doi:10.1207/s15324834basp2701
- Bernard, P., Gervais, S. J., Allen, J., Campomizzi, S., & Klein, O. (2012). Integrating sexual objectification with object versus person recognition: the sexualized-body-inversion hypothesis. *Psychological science*, 23(5), 469–71. doi:10.1177/0956797611434748
- Bruneau, E. G., Dufour, N., & Saxe, R. (2012). Social cognition in members of conflict groups: behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1589), 717–30. doi:10.1098/rstb.2011.0293
- Bruneau, E. G., Pluta, A., & Saxe, R. (2012). Distinct roles of the “shared pain” and “theory of mind” networks in processing others' emotional suffering. *Neuropsychologia*, 50(2), 219–31. doi:10.1016/j.neuropsychologia.2011.11.008
- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, 11(2), 49–57. doi:10.1016/j.tics.2006.11.004
- Carson, J. W., Keefe, F. J., Lynch, T. R., Carson, K. M., Goli, V., Fras, A. M., & Thorp, S. R. (2005). Loving-kindness meditation for chronic low back pain: results from a pilot trial. *Journal of holistic nursing : official journal of the American Holistic Nurses' Association*, 23(3), 287–304. doi:10.1177/0898010105277651
- Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2012). Decoding the Role of the Insula in Human Cognition: Functional Parcellation and Large-Scale Reverse Inference. *Cerebral cortex (New York, N.Y. : 1991)*. doi:10.1093/cercor/bhs065
- Coetzee, S. K., & Klopper, H. C. (2010). Compassion fatigue within nursing practice: a concept analysis. *Nursing health sciences*, 12(2), 235–243.
- Condon, P., & Barrett, L. F. (2013). Conceptualizing and experiencing compassion. *Emotion*.

- Condon, P., Desbordes, G., & Miller, W. (2013). Meditation increases compassionate responses to suffering. *Psychological Science*.
- Cox, C. L., Uddin, L. Q., Di Martino, A., Castellanos, F. X., Milham, M. P., & Kelly, C. (2012). The balance between feeling and knowing: affective and cognitive empathy are reflected in the brain's intrinsic functional dynamics. *Social cognitive and affective neuroscience*, 7(6), 727–37. doi:10.1093/scan/nsr051
- De Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual review of psychology*, 59, 279–300. doi:10.1146/annurev.psych.59.103006.093625
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3(2), 71–100. doi:10.1177/1534582304267187
- DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, 127(1), 1–56. doi:10.1093/qje/qjr050
- Desbordes, G., Negi, L. T., Pace, T. W. W., Wallace, B. A., Raison, C. L., & Schwartz, E. L. (2012). Effects of mindful-attention and compassion meditation training on amygdala response to emotional stimuli in an ordinary, non-meditative state. *Frontiers in Human Neuroscience*, 6(November), 1–15. doi:10.3389/fnhum.2012.00292
- Duerden, E. G., Arsalidou, M., Lee, M., & Taylor, M. J. (2013). Lateralization of affective processing in the insula. *NeuroImage*, 78C, 159–175. doi:10.1016/j.neuroimage.2013.04.014
- Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science (New York, N.Y.)*, 319(5870), 1687–8. doi:10.1126/science.1150952
- Dvash, J., Gilam, G., Ben-Ze'ev, A., Hendler, T., & Shamay-Tsoory, S. G. (2010). The envious brain: the neural basis of social comparison. *Human brain mapping*, 31(11), 1741–50. doi:10.1002/hbm.20972
- Eisenberg, N., Fabes, R., & Miller, P. (1989). Relation of Sympathy and Personal Distress to Prosocial Behavior : A Multimethod Study. *Journal of personality ...*, 57(1), 55–66.
- Engen, H. G., & Singer, T. (2012). Empathy circuits. *Current Opinion in Neurobiology*. doi:10.1016/j.conb.2012.11.003
- Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience and biobehavioral reviews*, 35(3), 903–11. doi:10.1016/j.neubiorev.2010.10.009
- Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., ... Kainbacher, M. (2011). The bystander-effect: a meta-analytic review on bystander

- intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137(4), 517–537.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–4. doi:10.1016/j.neuron.2006.05.001
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501.
- Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: an evolutionary analysis and empirical review. *Psychological bulletin*, 136(3), 351–74. doi:10.1037/a0018807
- Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 35(1), 4–26. doi:10.1038/npp.2009.129
- Harbaugh, W., Mayr, U., & Burghart, D. (2007). Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations. *Science*, 316, 1622–1624. doi:10.1126/science.1140738
- Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(2), 583–90. doi:10.1523/JNEUROSCI.4089-09.2010
- Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Social cognitive and affective neuroscience*, 2(1), 45–51. doi:10.1093/scan/nsl037
- Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and social psychology review*. doi:10.1207/s15327957pspr1003
- Hein, G., Silani, G., Preuschhoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68(1), 149–60. doi:10.1016/j.neuron.2010.09.003
- Herf, J. (2006). *The Jewish Enemy: Nazi Ideology and Propaganda During World War II and the Holocaust*. Harvard University Press.
- Izuma, K., Saito, D. N., & Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of cognitive neuroscience*, 22(4), 621–31. doi:10.1162/jocn.2009.21228
- Kim, J.-W., Kim, S.-E., Kim, J.-J., Jeong, B., Park, C.-H., Son, A. R., ... Ki, S. W. (2009). Compassionate attitude towards others' suffering activates the mesolimbic neural system. *Neuropsychologia*, 47(10), 2073–81. doi:10.1016/j.neuropsychologia.2009.03.017

- Klimecki, O. M., Leiberg, S., Lamm, C., & Singer, T. (2012). Functional Neural Plasticity and Associated Changes in Positive Affect After Compassion Training. *Cerebral cortex*, 1–10. doi:10.1093/cercor/bhs142
- Klimecki, O., & Singer, T. (2012). Empathic Distress Fatigue Rather Than Compassion Fatigue? Integrating Findings from Empathy Research in Psychology and Social Neuroscience. In B. Oakley, A. Knafo, G. Madhavan, & D. S. Wilson (Eds.), *Pathological altruism* (pp. 368–383). New York: Oxford University Press.
- Kok, B. E., Coffey, K. a, Cohn, M. a, Catalino, L. I., Vacharkulksemsuk, T., Algoe, S. B., ... Fredrickson, B. L. (2013). How Positive Emotions Build Physical Health: Perceived Positive Social Connections Account for the Upward Spiral Between Positive Emotions and Vagal Tone. *Psychological science*. doi:10.1177/0956797612470827
- Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. L. (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), 5648–53. doi:10.1073/pnas.1207992110
- Krienen, F. M., Tu, P.-C., & Buckner, R. L. (2010). Clan Mentality: Evidence That the Medial Prefrontal Cortex Responds to Close Others. *Journal of Neuroscience*, 30(41), 13906–13915. doi:10.1523/JNEUROSCI.2180-10.2010
- Lamm, C., Decety, J., & Singer, T. (2010). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage*, 54(3), 2492–502. doi:10.1016/j.neuroimage.2010.10.014
- Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-Term Compassion Training Increases Prosocial Behavior in a Newly Developed Prosocial Game. *PloS one*, 6(3), e17798.
- Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual review of psychology*, 58, 259–89. doi:10.1146/annurev.psych.58.110405.085654
- Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). *Face and Gesture 2011*, 298–305. doi:10.1109/FG.2011.5771414
- Lutz, A., Brefczynski-Lewis, J., Johnstone, T., & Davidson, R. J. (2008). Regulation of the neural circuitry of emotion by compassion meditation: effects of meditative expertise. *PloS one*, 3(3), e1897. doi:10.1371/journal.pone.0001897
- Mascaro, J., & Rilling, J. (2012). Compassion Meditation Enhances Empathic Accuracy and Related Neural Activity. *Social Cognitive and ...*
- Masten, C. L., Morelli, S. a, & Eisenberger, N. I. (2011). An fMRI investigation of empathy for “social pain” and subsequent prosocial behavior. *NeuroImage*, 55(1), 381–8. doi:10.1016/j.neuroimage.2010.11.060

- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behavior research methods instruments computers a journal of the Psychonomic Society Inc*, 33(4), 517–523.
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50(4), 655–63. doi:10.1016/j.neuron.2006.03.040
- Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42), 15623–8. doi:10.1073/pnas.0604475103
- Najjar, N., Davis, L. W., Beck-Coon, K., & Carney Doebbeling, C. (2009). Compassion fatigue: a review of the research to date and relevance to cancer-care providers. *Journal of Health Psychology*, 14(2), 267–277.
- Pace, T. W. W., Negi, L. T., Sivilli, T. I., Issa, M. J., Cole, S. P., Adame, D. D., & Raison, C. L. (2010). Innate immune, neuroendocrine and behavioral responses to psychosocial stress do not predict subsequent compassion meditation practice time. *Psychoneuroendocrinology*, 35(2), 310–5. doi:10.1016/j.psyneuen.2009.06.008
- Raz, G., Winetraub, Y., Jacob, Y., Kinreich, S., Maron-Katz, A., Shaham, G., ... Hendler, T. (2012). Portraying emotions at their unfolding: a multilayered approach for probing dynamics of neural networks. *NeuroImage*, 60(2), 1448–61. doi:10.1016/j.neuroimage.2011.12.084
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature reviews. Neuroscience*, 11(4), 264–74. doi:10.1038/nrn2805
- Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in cognitive sciences*, 16(3), 147–56. doi:10.1016/j.tics.2012.01.005
- Rudolph, U., Roesch, S., Greitemeyer, T., & Weiner, B. (2004). A meta-analytic review of help giving and aggression from an attributional perspective: Contributions to a general theory of motivation. *Cognition & Emotion*, 18(6), 815–848. doi:10.1080/02699930341000248
- Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5), 805–19.
- Salzberg, S. (2002). *Loving-Kindness: the revolutionary art of happiness*. Boston, MA: Shambhala Press.

- Sargeant, A., & Lee, S. (2004). Donor trust and relationship commitment in the U.K. charity sector: the impact on behavior. *Nonprofit And Voluntary Sector Quarterly*, *33*(2), 185–202.
- Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, *45*(13), 3054–67. doi:10.1016/j.neuropsychologia.2007.05.021
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain : a journal of neurology*, *132*(Pt 3), 617–27. doi:10.1093/brain/awn279
- Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*, 81–96. doi:10.1111/j.1749-6632.2009.04418.x
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, *48*(3), 564–84. doi:10.1016/j.neuroimage.2009.06.009
- van't Wout, M., & Sanfey, a G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. doi:10.1016/j.cognition.2008.07.002
- Vollhardt, J. R., & Staub, E. (2011). Inclusive altruism born of suffering: the relationship between adversity and prosocial attitudes and behavior toward disadvantaged outgroups. *The American journal of orthopsychiatry*, *81*(3), 307–15. doi:10.1111/j.1939-0025.2011.01099.x
- Weng, H. Y., Fox, a. S., Shackman, a. J., Stodola, D. E., Caldwell, J. Z. K., Olson, M. C., ... Davidson, R. J. (2013). Compassion Training Alters Altruism and Neural Responses to Suffering. *Psychological Science*. doi:10.1177/0956797612469537
- Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(49), 19761–6. doi:10.1073/pnas.1112324108
- Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience*, *15*(5), 675–80. doi:10.1038/nn.3085
- Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(27), 11382–7. doi:10.1073/pnas.0902666106

¹ While Roy et al (2012) use the term “affective meaning,” we use the term “emotional meaning” to emphasize this network’s role in constructing conceptually informed, elaborated emotions.