Accepted Manuscript

Title: How to think about emotion and morality: Circles, not arrows

Author: Kurt Gray Chelsea Schein C. Daryl Cameron

PII: S2352-250X(16)30230-5
DOI: http://dx.doi.org/doi:10.1016/j.copsyc.2017.06.011
Reference: COPSYC 476

To appear in:

Accepted date: 14-6-2017

Please cite this article as: K. Gray, C. Schein, C.D. Cameron, How to think about emotion and morality: Circles, not arrows, COPSYC (2017), http://dx.doi.org/10.1016/j.copsyc.2017.06.011

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
How to think about emotion and morality: Circles, not arrows.

Kurt Gray,¹ Chelsea Schein,¹ and C. Daryl Cameron²

¹University of North Carolina, Chapel Hill
²Pennsylvania State University

Word count: 2,331

Corresponding author:
Kurt Gray
University of North Carolina
CB 3270
Chapel Hill, NC 27599
kurtgray@unc.edu
Abstract

Emotion and morality are powerful conscious experiences. There are two ways to think about their psychological basis: arrows and circles. Arrows ground each experience in its own specialized mechanism (mechanism \( x \) causes phenomenon \( x \); mechanism \( y \) causes phenomenon \( y \)). Examples of arrows include when feelings of disgust are attributed to a specialized “disgust circuit” and when judgments of impurity are attributed to a specialized “purity foundation.” In contrast, circles—Venn diagrams—describe experiences as emerging from the overlap of more fundamental domain-general processes (different combinations of processes \( a, b, c \) cause both phenomena \( x \) and \( y \)). Circles are used by constructionist theories of emotion and morality, including the Theory of Dyadic Morality, which grounds moral judgment in the combination of norm violations, negative affect, and perceived harm. Despite the intuitive popularity of arrows, we show that scientific evidence is more consistent with circles.

Abstract word counts: 143
How to think about emotion and morality: Circles, not arrows.

Scientists often think about emotion, and about morality; however, they may be thinking about them incorrectly. Theories about the psychological basis of emotion and morality typically use “arrows,” which suggests that each variety of psychological phenomena is caused by its own special mechanism. In contrast, we suggest that psychological experience is better captured by overlapping “circles,” which suggests that conscious experiences arise from interactions between more general processes.

To illustrate these competing geometric metaphors, consider competing accounts of evil. Many people hold the intuitive belief that evil deeds are caused by the presence of evil within people’s souls (Darley, 1992; Miller, Gordon, & Buddie, 1999). For example, when non-social psychologists explain the Holocaust (Goldhagen, 1997), they often invoke the inherent evilness of the perpetrators. These explanations capture the “arrow” because an effect (an evil deed) arises from a single and qualitatively-similar cause (an evil soul). Arrows rest upon the assumption of “isomorphism,” which entails that cause and effect have 1) a similarity of form and 2) a unique one-to-one association. Formally, isomorphic arrows suggest that phenomenon $x$ is caused by mechanism $x$, and phenomenon $y$ is caused by mechanism $y$.

In contrast to “arrow” explanations, “circle” explanations reject isomorphism and instead ground phenomena in the overlap of more general ingredients (think Venn diagrams). Central to the idea of overlapping circles is emergence: phenomena arise from the combination of simpler elements, which together make more than the sum of their parts (Barrett, 2013; Lewontin, 2001). The same elements can combine to make different phenomenon because each can have different varieties and be present in different amounts. Formally, emergent circles suggest that
phenomenon $x$ is caused by the combination of elements $a, b_1, c_2$, and phenomenon $y$ is caused by the combination of elements $a, b_2, c_1$.

To explain evil deeds, a “circle” explanation would involve the combination of conformity, stress, ambiguity, and dehumanization, rather than the evil souls of perpetrators. These emergent situationist explanations may be intuitively unsatisfying to laypeople because they seem to “explain away” the phenomena of interest (i.e., evil; Miller et al., 1999), but they are commonly accepted in social psychology—except within emotion and morality. Here, we reveal how findings in emotion and morality are better captured by the idea of circles.

**Arrows of Emotion and Morality**

In psychology, intuitions of isomorphism lead scientists to explain emotions and morality with an arrow drawn from an eponymous mechanism. One popular theory of emotion—basic emotion—argues that we have six distinct feelings (e.g., fear, anger, disgust, joy, surprise and sadness) which are caused by six isomorphic mechanisms—one for each emotion (a fear mechanism, an anger mechanism, etc.; Ekman & Friesen, 1971; Ekman, 1992). See Figure 1. These mechanisms are thought to have evolved independently, to engender distinct experiences, and to have unique physiological and neurological signatures. Basic emotion is perhaps best captured by the animated movie *Inside Out*, in which each emotion is caused by its own little homunculus within the mind. Of course, this is a depiction for children, but its depiction of isomorphism was endorsed by eminent basic emotion scientists (Keltner & Ekman, 2015).

In morality, moral foundations theory (MFT) advocates for arrows (Haidt & Joseph, 2007), proposing that 5 distinct moral concerns (harm, fairness, loyalty, authority and purity) are caused by 5 distinct moral mechanisms (e.g., a harm module, a loyalty module; Haidt & Joseph, 2007). See Figure 1. In this isomorphic framework, each moral concern is also thought to be
linked to a specific basic emotion (e.g., disgust and purity; Horberg, Oveis, Keltner, & Cohen, 2009; Rozin, Lowery, Imada, & Haidt, 1999).

**Figure 1.** Isomorphic arrows in emotion (left) and morality (right), as described by basic emotions theory and moral foundations theory. With arrows, psychological experiences are tied to the operation of an eponymous module.

Despite the popularity of these theories, there is little evidence for isomorphic arrows in either emotions or morality. Emotions are neither consistent nor specific across situations, arguing against one-to-one mechanisms (Barrett et al., 2007; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). Even the same emotion can be experienced very differently; anger can be the white heat of road rage or cold brittleness toward an obnoxious boss (for similar examples, see Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). The lack of isomorphism is also revealed by physiology: discrete emotions lack anatomical specificity in the brain (Kober et al., 2008; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012) and body (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Kreibig, 2010; Mauss & Robinson, 2009).

Like emotions, moral concerns cannot be captured by isomorphic arrows: moral concerns largely overlap with each other (loyalty and authority, $r = .88$, Graham et al., 2011; purity and harm, $r = .87$, Gray & Keeney, 2015b), and all are tied to intuitive perceptions of harm (Gray,
Schein, & Ward, 2014; DeScioli, Gilbert, & Kurzban, 2012; Schein & Gray, 2015). To be sure, there are differences in moral content—bizarre sex acts are not the same as rigging an election—but these differences in content (e.g., “purity” versus “fairness”) need not require distinct moral mechanisms. Instead, diversity in moral content can be captured by general dimensions including severity and weirdness (Gray & Keeney, 2015b, 2015a)—and by cultural conceptions of harm (Shweder, 2012).

Historical (Turiel, Hildebrandt, Wainryb, & Saltzstein, 1991) and recent work (Schein, Goranson, & Gray, 2015; Schein, Hester, & Gray, 2016) suggests that differences in moral judgment across people are best understood as differences in how harm is perceived. For example, conservatives (but not liberals) view violations of chastity and patriotism as harmful, and hence they also view them as immoral (Schein, Ritter, & Gray, 2016). Rather than distinct moral mechanisms, we suggest that MFT provides a taxonomy of politically-variable values—similar to Schwartz’s (Schwartz & Bilsky, 1987) or Janoff-Bulman’s models (Janoff-Bulman & Carnes, 2013). With little evidence for isomorphic arrows, we suggest that circles are a better analogy for psychological experience.

Circles

In contrast to isomorphic arrows of basic emotions and MFT, constructionist theories argue that psychological experiences emerge from the combination of fundamental ingredients (Barrett & Russell, 2014). A popular analogy for constructionism is baking (Barrett, 2009). There are a set of basic ingredients—flour, water, eggs, baking powder, baking soda, salt, sugar, butter—out of which you can make a dazzling array of different baked goods, including croissants, scones, and cookies. One does not need a special cookie mechanism (i.e., an isomorphic arrow) to make a cookie instead of a croissant. Instead, one simply combines the
ingredients differently. Focusing on what makes a cookie essentially different from a croissant misses the point: the real scientific focus should be on how the underlying mechanisms combine in different ways.

Consistent with constructionism, research reveals that different emotions emerge from the combination of two fundamental ingredients (i.e., the overlap of two circles)—core affect and conceptualization (Cameron, Lindquist, & Gray, 2015; Lindquist, 2013; Barrett, 2012, 2013). Core affect is the physiological state of the body, and consists of two dimensions: valence (positive/negative) and arousal (high/low; Russell, 2003). Conceptualization involves making sense of that physiological state and includes considerations of situational context, cultural knowledge and past experiences. When core affect and conceptualization combine, the result is discrete emotions (Figure 2; for more discussion, see Barrett, 2013). For example, seeing a snake on a plane during turbulence constructs a feeling we call “fear”. Seeing a snake during a summer day at the zoo constructs a feeling we call “interest”. Understanding emotions as overlapping circles gives both nuance and complexity to emotional life; rather than only 6 emotions, there is an infinity of experiences, depending upon people’s exact situation and bodily state.

Emergence through combination also occurs with morality. As depicted in Figure 2, the constructionist Theory of Dyadic Morality (Gray, Young, & Waytz, 2012; Schein & Gray, in press, 2015) suggests that moral condemnation emerges from the overlap of three elements (i.e., circles): (negative) core affect (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), a (culturally determined) norm violation (Nichols, 2002) and (perceptions of) harm. The presence or potential for harm is what separates negative norm violations (e.g., spitting in your soup at a restaurant) from immoral actions (e.g., spitting in someone else soup; Turiel, 1983; Smetana, 1985; Huebner, Lee, & Hauser, 2010). Importantly, the harm central to moral cognition is not
objective and reasoned, but is instead intuitive and perceived. This is why objectively harmless acts (Haidt, 2001) still engender moral condemnation when they feel harmful, like pretending to shoot someone in the head (Cushman, Gray, Gaffey, & Mendes, 2012), or having safe sex with your sibling (Royzman, Kim, & Leeman, 2015).

**Figure 2.** Circles in emotion (left) and morality (right). Emotion constructionism grounds emotion (E) in the combination of core affect, and elements of conceptualization (e.g., cultural knowledge and situational context). The Theory of Dyadic Morality grounds perceived immorality (I) in the combination of a norm violation, (negative) core affect, and perceived harm.

Harm itself stems from three ingredients: 1) an intentional agent, 2) causing damage to 3) a vulnerable patient. Together, these three ingredients are represented by two interacting minds—a perpetrator and a victim—which is why this theory is called *dyadic* morality (Gray & Wegner, 2011b; Wegner & Gray, 2016). Reflecting the importance of each of these features, acts are more immoral when they are more intentional (murder vs. manslaughter; Cushman, Young, & Hauser, 2006; DeScioli, Asao, & Kurzban, 2012; Gray, 2012), cause more perceived suffering (murder vs. attempted murder; Schein & Gray, 2015) and have a clear causal
connection between perpetrator and victim (physical abuse vs. negligence; Malle, Guglielmo, & Monroe, 2014; Spellman & Gilbert, 2014; DeScioli, Christner, & Kurzban, 2011). These features can be represented by overlapping circles, or alternatively, the technique of “theory mapping” places fundamental elements within upward facing “)” (Gray, in press). Figure 3 illustrates that that negative affect, norm violation and harm combine to make immorality, and harm is itself made up of agent, patient and causation.

![Figure 3. Overlapping circles as depicted by “theory mapping” (Gray, 2016). Each element within a “)” symbol is a circle whose collective overlap constructs the element to which it points.](image)

For example, perceived harm is made up of the combination of intentional agent, the causation of damage, and a vulnerable patient (i.e., the moral dyad).

With overlapping circles, there is no need for a specific moral faculty—and especially no need for a moral faculty for each kind of cultural concern. Rather, diverse moral concerns arise through cultural variation in values (i.e., norms) and perceived harm (Rachels, 1986; Shweder, 2012). For example, Brahmin Indians think that failing to follow funeral rites harms the immortal soul, explaining why they moralize this “purity” concern (Shweder, 2012).

Substantial research supports the overlapping circles of constructionist theories of emotions and morality (Cameron et al., 2015; Cheng, Ottati, & Price, 2013; Gray & Wegner,
Of course, other research argues for distinct isomorphic mechanisms (e.g., purity is specially tied to a disgust mechanism; Rozin et al., 1999; Scott, Inbar, & Rozin, 2016), but this research often fails to establish discriminant validity by including appropriate control conditions or analyses (for discussions, see Cameron et al., 2015; Gray & Schein, 2016). When studies do contain appropriate controls and analyses, there is no evidence for isomorphic mechanisms (Cheng et al., 2013).

It's the Real Thing

Just as laypeople often reject emergent situationist explanations of evil, advocates of isomorphism reject constructionist claim because they believe it “explains away” emotions and morality (for discussion, see Barrett, 2012, 2015). Nothing could be further from the truth; constructionism embraces the existence of many emotions and moral concerns, each of which feels very special. Indeed, while basic emotion and MFT forces experiences into a small set of researcher-defined taxonomies, constructionism allows for as many shades of emotion and morality as there are different cultures and situations. By arguing against the idea of $n$ conscious feelings stemming from $n$ specialized psychological mechanisms, we suggest that constructionism better embraces the variability of human experience.

As an analogy for how emotions and morality can be “real” without isomorphism, consider the reality of Coca-Cola. This beverage certainly has a unique taste, and a unique identity from other beverages like Pepsi. However, the difference between Coke and Pepsi does not stem from a special and distinct “Coke” or “Pepsi” production process. Instead, both share the same set of basic ingredients, and the differences lie with how these ingredients are all combined—and the branding (conceptualization) layered on top of those ingredients.
The 1942 slogan of coke was “The only thing like Coca-Cola is Coca-Cola itself,” but at the end of the day, this special beverage is a variety of flavored, carbonated, sugar water—made meaningful by packaging and sentimental commercials. Likewise, ‘the only thing like fear or disloyalty may be fear or disloyalty itself,’ but that doesn’t mean there is an underlying essence—a special production process—for these experiences. Instead, these conscious experiences arise from combinations of more fundamental psychological ingredients made meaningful by the “branding” of cultures and cognitions.

**Conclusion**

The power and diversity of emotion and morality is obvious, but how exactly we should think about these phenomena is less obvious. Despite the intuitive appeal of isomorphic arrows, there is relatively little evidence for 5 or 6 special emotion or morality mechanisms each tied to 5 or 6 different experiences. Instead, emotion and morality is better understood through the constructionist analogy of overlapping circles, in which experiences emerge from the combination of more fundamental ingredients. We admit that ideas of emergence may be less intuitive than those of isomorphism, but psychologists have long known that our intuitions lead us astray (Ross & Ward, 1997). Importantly, faulty intuitions are not insurmountable: if we can embrace emergent situationist accounts of social behavior (e.g., explaining why people do evil), we can also embrace emergent accounts of psychological experiences. Emotions and morality are too important to our everyday lives—and to our science—to not think clearly about them.
References


This concluding chapter provides a generative framework for psychological constructionism with an eye toward future research.


Adapting a constructionist perspective, this paper overviews research at the intersection of morality and emotions. It finds little support for specific emotion-morality links.


  *Theory is paramount in science. This paper offers a new tool for psychologists to visualize theory.*


This paper overviews the history of psychological constructionism and then overviews the psychological constructionist model of emotion.


This paper provides a comprehensive overview of the Theory of Dyadic Morality. It looks at both the content and mechanisms underlying moral cognition.


**Highlights**

There are two frameworks for understanding conscious experience: arrows and circles. Arrows link one special mechanism to one unique mental experience. Circles explain mental experience as the overlap of more basic processes. Basic emotions and moral foundations advocate for arrows, constructionism for circles. We suggest that circles are a better way to think about emotion and morality.