AMERICAN PSYCHOLOGICAL ASSOCIATION

Emotion

ISSN: 1528-3542

2023, Vol. 23, No. 1, 182–193
https://doi.org/10.1037/emo0000981

Emotional Change: Neural Mechanisms Based on Semantic Pointers

Paul Thagard[1], Laurette Larocque[2], and Ivana Kajić[3]
[1] Department of Philosophy, University of Waterloo
[2] Guelph, Ontario, Canada
[3] Department of Computer Science, University of Waterloo

Emotional change includes generation of new emotions, switching from one emotion to another, and alteration of the frequency and intensity of emotions. Psychotherapists help clients to reduce negative emotions such as sadness and anxiety and increase positive emotions such as happiness and hope. We explain such emotional shifts by the semantic pointer theory of emotions, which views them as brain processes that integrate neural representations of situations, appraisals of the goal-relevance of those situations, and physiological reactions to the situations. This theory can explain many kinds of emotional change, including the generation and shifting of mixed, nested, and dispositional emotions.

Keywords: embodiment, emotion, emotional change, psychotherapy, semantic pointers

Supplemental materials: https://doi.org/10.1037/emo0000981.supp

People go to psychotherapists in search of emotional change because they are afflicted with negative emotions such as sadness, anxiety, fear, and shame. Effective therapists can help them to replace these emotions with positive ones such as happiness, confidence, serenity, and pride. We are not offering a general theory of psychotherapy but using it as an important illustration of emotional change.

Emotional change is common in everyday life. A toddler can go in seconds from happily playing to having an angry tantrum when another child interferes. A professor can be enjoying the day until a grant rejection arrives, producing a transition from contentment to negative emotions such as disappointment and frustration.

An acceptable theory of emotion needs to be able to explain emotional change as well as many other phenomena such as the contribution of emotions to action. There is currently no generally accepted theory of emotion, as debate continues on whether emotions are best understood as innate brain programs, cognitive appraisals, bodily processes, or social constructions. We contend that these theories provide perspectives on emotional change that should be viewed as complementary rather than competing.

We offer a new explanation of emotional change based on the semantic pointer theory of emotions (Kajić et al., 2019; Thagard, 2019a, 2019b; Thagard & Schröder, 2014). Semantic pointers provide a new way of thinking about how the brain can combine perception of bodily changes with cognitive judgments. Different combinations lead to different emotions and to changes from one pattern of emotions to another.

We summarize how prominent theories of emotion can explain varieties of emotional change. Emotional change looks very different from the perspectives of innate programs, appraisals, embodiment, and construction, but the insights of these perspectives can be amalgamated into a sufficiently rich theory of neural mechanisms. We show how the semantic pointer theory of emotions accomplishes this amalgamation and applies to the varieties of emotional change. Hence this theory deserves consideration as an alternative to the more familiar theories.

We might be expected to begin with a definition of the concept of emotion, but an approach more compatible with psychological research views concepts as combining exemplars (standard examples), typical features, and explanations (Blouw et al., 2016; Murphy, 2002). Accordingly, Table 1 offers a "three-analysis" of *emotion* that specifies exemplars such as happiness, features such as physiology and appraisal, and explanations concerning experiences and behaviors.

Here are two cases of emotional change in psychotherapy from Dr. Larocque's independent practice, modified to maintain anonymity. Steve is a man in his 20s who experienced sexual abuse as an adolescent that induced prolonged negative emotions including shame, embarrassment, disgust, bitterness, and helplessness. Therapy helped him to experience less shame and increased frequency of positive emotions such as happiness and hope. Joan is a woman in her 40s who sought help for feelings of anxiety, depression, and guilt about her weaknesses in doing her job and looking after her family. Therapy helped her to be less anxious and more confident about dealing with work and family responsibilities. Theories of emotion should be able to explain such shifts.

Computer simulations are valuable in cognitive science because they provide detailed models of explanatory mechanisms whose performance can be compared with human performance (Thagard, 2012). Accordingly, we have approximated the Steve and Joan cases

This article was published Online First November 29, 2021.
Paul Thagard https://orcid.org/0000-0002-6594-4761
Ivana Kajić https://orcid.org/0000-0001-8992-2391
The section of the article on moods is adapted from Thagard's *Psychology Today* blog, *Hot Thought*.
Correspondence concerning this article should be addressed to Paul Thagard, Department of Philosophy, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. Email: pthagard@uwaterloo.ca

Table 1
Three-Analysis of Emotion

Exemplars	Happiness, sadness, fear, anger, disgust, surprise, shame, embarrassment, pride, etc.
Typical features	Physiological changes, cognitive appraisals, social influences including linguistic ones in humans, neural patterns in multiple brain areas, enjoyable or painful experiences, varying intensity, resulting actions.
Explanations	Explains: experiences, behaviors, reports, motivations, communications. Explained by: molecular, neural, mental, and social mechanisms.

Note. From *Brain-Mind: From Neurons to Consciousness and Creativity* (p. 135), by P. Thagard, 2019, Oxford University Press. Copyright 2019 by Oxford University Press. Reprinted with permission.

using the POEM ("POinters-EMotion") computational model of emotions as semantic pointers. Because of length limitations, the Joan simulation is reported in online supplemental material.

Theories of Emotion

Adequately surveying current theories of emotions requires a handbook or textbook, (e.g., Barrett et al., 2016; Keltner et al., 2018). More concisely, we summarize the implications of currently prominent emotion theories in Boxes 1–5. Our focus in this article is on emotional change, but an acceptable theory of emotion has to provide answers to many other important questions (Supplement 1 in the online supplemental materials).

Appraisal Theories

- **Theory of emotion:** Emotions are cognitive appraisals of the relevance of a situation to goals.
- **Advocates:** Nussbaum (2001); Oatley (1992); Ortony et al. (1988); Scherer et al. (2001).
- **Hypothesis about change:** Emotions begin and change because of new appraisals of situations.
- **Mechanisms:** Inferences about consequences; parallel constraint satisfaction.
- **Computational models:** Marsella and Gratch (2014); Thagard and Aubie (2008).

Physiological Theories

- **Theory of emotion:** Emotions result from physiological reactions to events.
- **Advocates:** Damasio (1994); James (1884); Prinz (2004).
- **Hypothesis about change:** Emotions begin and change because of physiological changes.
- **Mechanisms:** Somatic markers.
- **Computational models:** Wagar and Thagard (2004).

Basic Emotions Theory

- **Theory of emotion:** Emotion is an automatic appraisal influenced by evolution and personal history.
- **Advocates:** Ekman (2003); Ekman and Cordaro (2011).
- **Hypothesis about change:** Emotions begin and change because of programmed appraisals.
- **Mechanisms:** Affect programs—evolved instructions in the brain.
- **Computational models:** ?

Theory of Constructed Emotion

- **Theory of emotion:** Emotions are constructed by the brain based on culture and physical properties of the body.
- **Advocates:** Barrett (2017); Harré (1989).
- **Hypothesis about change:** Emotions begin and change because of new constructions.
- **Mechanisms:** Predictive coding
- **Computational models:** ?

The Semantic Pointer Theory of Emotions

- **Theory of emotion:** Emotions are brain processes based on semantic pointers.
- **Advocates:** Kajić et al. (2019); Thagard (2019a; 2019b); Thagard and Schröder (2014).
- **Hypothesis about change:** Emotions begin and change because of the formation of semantic pointers and competition among them.
- **Mechanisms:** Firing patterns in groups of neurons, binding into semantic pointers, competition between semantic pointers.
- **Computational models:** Kajić et al. (2019).

Each theory of emotion comes with its own account of emotional change. If emotions are fundamentally cognitive appraisals, then changing emotions is reappraisal that occurs because of different situations, different goals, or different inferences about the relevance of situations to goals. In contrast, if emotions are responses to physiological changes, then changing emotions comes from changes in physiology, for example from varying patterns of bodily states such as heart rate. The theory that ties emotions to innate affect programs sees change as resulting from the running of different programs. Theories that view emotions as resulting from social or psychological constructions explain change as resulting from the generation of different constructions. There are also hybrid theories that view emotions as combinations of cognitions and feelings (Reisenzein, 2017). Finally, our theory of emotions as semantic pointers provides explanations of changes resulting from neural mechanisms described below.

Following current philosophy of science, we take a psychological or biological theory to be a description of mechanisms, which are combinations of connected parts whose interactions produce regular changes (Bechtel, 2008; Craver & Darden, 2013; Thagard, 2019c). For a psychological theory of emotional change, the regular changes that result from parts (mental representations) and interactions (associations, inferences) concern the starting, stopping, switching, and mixing of emotions. In brain mechanisms, the parts are neurons and the primary interactions are excitation and inhibition resulting from synaptic connections.

There is a natural translation between mechanisms and computer models. The parts in a mechanism can be described by data structures in a computer program, and the interactions can be specified by algorithms. Then running the computer program helps to show that the parts and their interactions yield the expected results—the regular changes to be explained. Computer models make two major contributions to psychological theory: requiring that the mechanisms they posit be rigorously specified and helping to determine whether the mechanisms can generate the empirical phenomena to be explained. Computational models of emotion are reviewed by Hudlicka (2011), Marsella and Gratch (2016), and Reisenzein et al. (2013).

The Semantic Pointer Theory of Emotions

Theories of emotions have been around since the ancient Greeks, but the new semantic pointer theory uses neural mechanisms to incorporate central ideas from preceding theories, including both appraisal and physiology. We now describe the mechanisms proposed by the semantic pointer theory and informally characterize the computational model that implements these mechanisms.

Mechanisms

Table 2 summarizes the three main mechanisms applied by semantic pointer theories to numerous kinds of thinking: neural firing patterns, binding of neural patterns into richer patterns called semantic pointers, and competition among the semantic pointers (Crawford et al., 2016; Eliasmith, 2013; Thagard, 2019a). Neurons are brain cells that receive excitatory and inhibitory inputs from other neurons, building up positive charge. If a neuron's charge passes a threshold, then it fires, sending electrical signals that excite or inhibit other neurons.

The firing pattern of a single neuron is the temporal sequence of firing, for example FIRE REST FIRE FIRE REST. What matters is not just how fast the neuron is firing but also the particular sequence of firing or not firing over time. In principle, a single neuron's firing pattern could stand for a detected stimulus such as the sound of a trumpet, but the brain usually employs numerous neurons to represent stimuli. A neural representation is a group of neurons, each with its own firing pattern, that collectively have a firing pattern that stands for features of the world such as the sound or look of a trumpet.

For more complex representations such as *yellow trumpet*, the patterns of neural firing for *yellow* and for *trumpet* must somehow be bound together. A simple way is by synchronizing the firing of neurons in different groups, but such synchrony does not yield patterns that can be subject to further binding, for example in the sentence "The yellow trumpet is on top of the black table." A more powerful means of binding different patterns of neural firing is *convolution,* which weaves together firing patterns into a new ones that can be subject to further binding; for mathematical details, see Plate (2003) and Eliasmith (2013).

Table 2
Neural Mechanisms of Cognition and Emotion

Parts	Interactions	Changes
Neurons	Excitation and inhibition	Firing patterns
Neural groups	Binding by convolution	Semantic pointers
Semantic pointers	Competition by inhibition	Perceptions, emotions

Convolution can be used to bind patterns of firing into especially rich patterns that Eliasmith (2013) calls *semantic pointers*. These new patterns are special in that they weave together patterns of firing into new patterns that can function as symbols in further inferences, while allowing reconstitution of the sensory inputs that were captured by the patterns that were bound together. For example, the inputs that led to formation of the neural representation *yellow* in *yellow trumpet* can be partially reclaimed from the firing pattern of the newly formed *yellow trumpet*. Hence semantic pointers can function as a combination of both symbolic and sensory information.

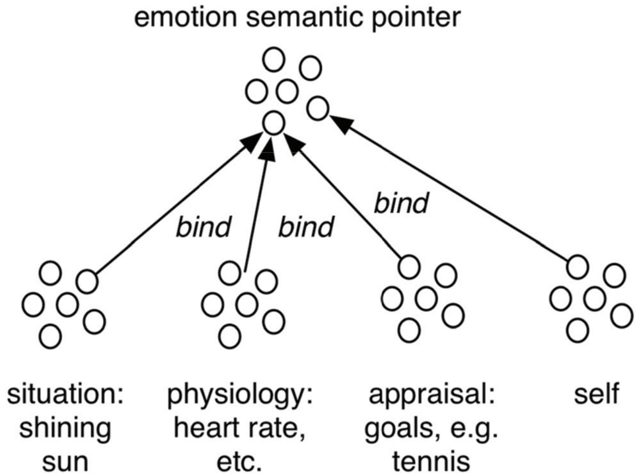
This combination makes semantic pointers powerful as an account of how the brain produces emotions. Your feeling happy that the sun is shining requires representation of the situation that the sun is shining, which could be purely verbal but more likely includes sensory images of the bright sun and of the warmth on your skin. It might also include representation of you as the possessor of the emotion.

Most problematically, feeling happy requires representation of the positive response that you have toward the situation. According to appraisal theories, this response comes from inferences concerning the extent to which the situation furthers your goals, for example if the shining sun means you can play tennis. According to physiological theories, the happy response comes from the brain's detecting changes in bodily states such as breathing rates.

The semantic pointer theory of emotions hypothesizes that emotions combine appraisal and physiology by convolution of patterns of neural firing. Appraisal generates one neural pattern, and physiology generates another neural pattern, along with the neural patterns for the situation and the self. All of these can be bound together into a unifying neural pattern that has both the symbolic and perceptual aspects of semantic pointers. The symbolic aspect allows the representation of happiness to be used in verbal descriptions and explanations, for example when you say that you are smiling because you are happy that the sun is shining. The perceptual aspect allows the semantic pointer to compress some of the physiological changes that went into the emotion. Figure 1 depicts how the semantic pointer for being happy that the sun is shining binds the contributing neural patterns for the situation, the self, appraisal, and physiology.

How does a neural pattern represent the result of an appraisal? Most generally, appraisal of the extent to which a situation accomplishes a person's goals is a constraint satisfaction problem that can be computed by neural networks (Thagard & Aubie, 2008). The positive constraints for happiness are that a situation should satisfy goals such as pleasantness and predictability, and the negative constraints are that happiness is incompatible with conflicting emotions such as sadness. The simplest way for a neural network to represent the resulting computation would be to have a neuron HAPPY that fires rapidly when the constraints are satisfied. A more neurologically plausible way to get the same result is when the neural group for the semantic pointer corresponding to "I am happy that the sun is shining" has a rapid and distinctive pattern of firing.

Figure 1
Semantic Pointer for the Emotion of Being Happy That The Sun is Shining



Note. The circles indicate thousands or millions of neurons.

Appraisal does not always need to go through the full process of parallel constraint satisfaction, if an organism has linguistic knowledge that incorporates past experiences. Then the positive or negative emotional valences that are attached to words such as "shining sun" can combine to provide a fast and frugal calculation of the extent to which a situation satisfies goals (Kajić et al., 2019). Such uses of language incorporate cultural constructions into appraisal because the valences attached to words are largely the result of cultural experiences. For example, whether the word *gun* has positive or negative emotional value depends on a society's attitude toward violence. Culture can also enter into the more general process of parallel constraint satisfaction, because the goals with respect to which situations are evaluated are partly based on social values such as individual achievement versus group harmony.

How does a neural pattern represent physiological changes? The body has many internal sensors that send information to brain areas such as the amygdala and the insula. Neural groups in these areas aggregate that information by patterns of firing, which thereby serve as representations of the state of the body. As shown in Figure 1, semantic pointers integrate this bodily information with the result of appraisal, along with representations of the situation and the self, in organisms that have self-representations.

The third neural mechanism crucial for explaining emotion in terms of semantic pointers is competition, which is most easily understood in terms of two neurons. Suppose there were a neuron for HAPPY and another for SAD. Then competition between the two neurons is easily modeled by having the HAPPY neuron inhibit the SAD neuron, and vice versa. Then if one is firing rapidly, the other usually is not.

Competition is harder to understand when it is between the complex neural groups that correspond to semantic pointers, but an analogy helps. When two basketball players compete one-on-one, the winner is determined only by how well they each score. But competition at the team level results from many individual competitions among individual players, with result that one team wins and the other loses. Similarly, the competition between semantic pointers results from the operation of inhibition in many pairs of neurons included in the different neural groups. Semantic pointer competition overall emerges from inhibitory links between the neurons that make up the different neural groups, allowing the HAPPY semantic pointer to compete with the SAD semantic pointer.

Inhibition is a computational mechanism that can be interpreted at two different but related levels. One is where a single neuron inhibits the activity of another neuron and the other is where neural groups representing concepts inhibit each other. Although both mechanisms are used in POEM, the latter level is the primary driver for the competition mechanism when neural groups representing semantic pointers inhibit each other.

Computational Model

That these mechanisms can be rigorously specified to generate explanations of emotional phenomena is shown by the computer model POEM (Kajić et al., 2019). This model uses the neural simulator NENGO, which applies the computational principles of the Neural Engineering Framework of Eliasmith and Anderson (2003). POEM uses vectors as inputs to represent aspects of situations, and NENGO transforms these into neural patterns distributed over thousands of simulated neurons.

Similarly, physiological information is represented by vectors that are transformed into neural patterns. Binding of neural patterns into new ones is accomplished by the mathematical operation of convolution, defined as an operation on vectors and accomplished in NENGO by neural groups that generate new patterns of firing, including semantic pointers for emotions.

For fast and frugal appraisal, POEM uses linguistic information based on affect control theory (Heise, 2007). This information is derived from empirical studies that rate words on three dimensions: evaluation (goodness vs. badness), potency (powerfulness vs. powerlessness), and activity (liveliness vs. torpidity). Then linguistic descriptions captured by vectors can generate appraisals based on the ratings of words on these dimensions, for example with *sun* and *shining* generating positive appraisals because they rate highly on goodness, powerfulness, and liveliness. For a more general appraisal mechanism that applies to nonlinguistic animals, POEM could also implement the constraint satisfaction mechanism describe by Thagard and Aubie (2008). Unlike the Heise approach, which uses language associations as a stand-in for appraisal by goal evaluation, constraint satisfaction explicitly assesses goal satisfaction.

POEM implements competition between semantic pointers by means of numerous inhibitory links between the neurons in competing neural groups. The result is that the model can decide whether a situation is happy or sad by competition between the corresponding semantic pointers. POEM was originally used for simulation of six important aspects of emotions: (a) Some stimuli generate immediate emotional reactions; (b) Some emotional reactions depend on cognitive evaluations of external stimuli; (c) Bodily states influence the generation of emotions; (d) Some emotions depend on interactions between physiological inputs and cognitive appraisal; (e) Some emotional reactions concern syntactically complex representations; (f) Mixed emotions arise from ambivalent appraisals of events. New simulations of the Steve and Joan cases are described below and in the online supplemental material.

Explanations of Emotional Changes

Now we can consider how the semantic pointer theory of emotions fares in providing explanations for the most important phenomena of emotional change. These include the onset and cessation of simple, mixed, and nested emotions, along with switching between emotions and between patterns of emotions as in psychotherapy. Dispositional emotions and moods are also accommodated. Only some of these phenomena have been modeled computationally using POEM, but we can sketch broadly how emotional change can result from the three mechanisms of neural firing, binding into semantic pointers, and competition.

Onset and Cessation of Occurrent Emotions

The simplest kinds of emotional change that need to be explained are the onset and cessation of emotions. There are hundreds of different emotions signified by words in English and other languages. People can go from a state with no recognized emotion to pleasant states such as admiration, adoration, amazement, amusement, annoyance, awe, calmness, curiosity, excitement, humor, happiness, interest, intrigue, joy, love, pride, relief, satisfaction, serenity, and wonder. Or they can end up in unpleasant states such as anger, anxiety, boredom, confusion, craving, disgust, embarrassment, fear, horror, pain, sadness, and shock (Cowen & Keltner, 2017). Other emotions states such as desire, nostalgia, and surprise can be pleasant or unpleasant depending on the context and co-occurrence of other emotions. A theory of emotions should be able to explain the commencement of each of these emotions.

Because the semantic pointer theory of emotions understands them as brain processes, explanation of the starting and stopping of emotions is straightforward. The starting of an emotion requires the formation of several neural representations that can occur simultaneously, thanks to the parallel processing in the brain across numerous brain regions. The situation that the emotion is about is represented by firing patterns in verbal and/or sensory areas. The appraisal of the relevance of the situations to a person's goals results from interactions among cortical areas such as the dorsolateral prefrontal cortex and the orbital frontal cortex, and among subcortical areas such as the amygdala and nucleus accumbens. At the same time, physiological changes prompted by perception of the situation are detected by brain areas such as the insula and the amygdala. None of these representations by itself constitutes an emotion.

An unconscious emotion is formed when these three neural representations are bound together into a semantic pointer by convolution, with possible inclusion of a representation of the self. This is the onset of an emotion but is not yet experienced as such. For conscious experience, the semantic pointer created by binding representations of the situation, appraisal, and physiological change must outcompete other semantic pointers that have been formed by binding. Then the conscious experience of emotion begins. Emotional consciousness is then a special case of the general theory of consciousness as resulting from competition among semantic pointers (Thagard, 2019a; Thagard & Stewart, 2014). For example, Steve feels ashamed when he recalls incidents of sexual abuse and binds memories of the situations with both physiological changes such as a racing heart and appraisals such as the abuser threatening his goal to survive. Joan feels anxious when her brain binds thinking about the difficulties of balancing family and work with physiological changes such as increased stomach turmoil.

The cessation of an emotion merely requires removal of one or more of the ingredients that went into its onset. If the situation is no longer salient because of changes in verbal or perceptual information, then the semantic pointer begins to unravel. If you are happy that the sun is shining, but the day becomes cloudy, then the words and images for the shining sun dissipate unless you can keep them in memory through rumination. At the same time, the appraisal may change if different goals become active, for example if you realize that the shining sun could lead to skin cancer. Physiology can also alter dramatically, for example if thought of cancer increases your cortisol levels. Any of these changes can lead to the unraveling of the semantic pointer, or at least to its declining below the limit of consciousness. A therapist can help Steve to cease feeling shame about his abuse by modifying physiology (through techniques such as deep breathing and exercise) and by modifying appraisal (through emphasis that the abuse was not his fault and the threat is passed).

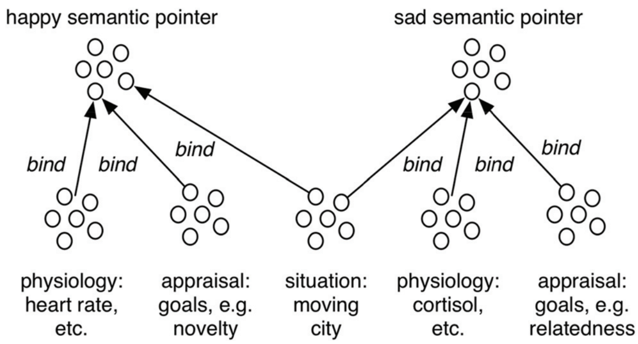
Emotion semantic pointers can also drop out of consciousness because they are outcompeted by other emotions, which is discussed later as switching. The computational model POEM has been used to simulate the generation of emotion semantic pointers, with the appraisal function performed linguistically rather than by the more general constraint satisfaction process.

Mixed and Nested Emotions

The most common kind of emotional change involves the onset of relatively simple emotions such as happiness and sadness. But there also more complex kinds of emotions that need to be accounted for as part of a full theory of emotional change. Mixed emotions occur when people feel pleasant and unpleasant emotions at the same time, for example the combination of pride and sadness that takes place when a child grows up and leaves home (Heavey et al., 2017). Joan has mixed emotions about her family because she loves them but is overwhelmed by the amount of care they require. How can people have more than one emotion at the same time?

From the perspective of semantic pointers, the answer is that the same situation can generate different appraisals and different physiological changes. How this works is most evident in the case where the mixed emotions have different valences, say happiness and sadness, as in Figure 2. Suppose the situation is that you are moving to a new city, which makes you happy that you are going to have new adventures, but sad that you are leaving your old life behind. The mixed emotion consists of the co-occurrence of two different semantic pointers, both of which include representations of the situation, but with different appraisals and physiology. The happiness appraisal comes from accomplishing goals of novelty, whereas the sadness appraisal comes from failure to accomplish goals of relatedness. Simultaneously, the physiological signals going to the amygdala and insula are a mixture of positive changes

Figure 2
Semantic Pointer for the Mixed Emotion of Being Both Happy and Sad About Moving



such as rapid heart rate, and negative one such as increased cortisol.

Ordinarily, competition among semantic pointers for happiness and sadness would result in one of them dominating, so how can the brain generate both happiness and sadness at the same time? Recall the basketball analogy for competition. Sometimes two teams end up tied at the end of a regular game because neither is able to dominate the other. Similarly, if the situation generates strong appraisals and physiological changes for two different emotions, then the semantic pointers end in a tie in which neither can outcompete the other. Mixed emotions result.

In POEM, the mixed appraisal of situations is modeled by showing that the same concept can elicit different emotions. The model predicts overwhelmingly positive emotions, such as contentment, happiness, and pleasure, when it is presented with the inputs "cake" and "taste." However, when the focus shifts from "taste" to "obesity," the model changes its output to a combination of emotions with negative and positive valence, such as love and sadness (see Simulation 6 in Kajić et al. [2019] for details).

Nested emotions, which are emotions about emotions such as fear of embarrassment, are more complicated. It is mysterious how nested emotions could work as mere appraisals or physiological changes, but semantic pointers have no problems with nested emotions because of their recursive character. The brain is capable not only of bindings, but also of bindings of bindings of bindings. A neural representation of embarrassment requires binding of representations of: (a) a situation such as giving a bad talk; (b) appraisal that this event is contrary to goals such as being respected by others; (c) physiological changes such as sweaty palms; and (d) the self. All these get bound together into a semantic pointer that represents being embarrassed about speaking badly, as illustrated in Figure 3.

Then this semantic pointer can be bound into a more complicated semantic pointer that includes the embarrassment of the situation, the appraisal that embarrassment is contrary to one's personal goals such as feeling good, and the physiological accompaniments of fear such as a pounding heart. So nested emotions consist of neural representations of representations of representations, all accomplished by semantic pointers.

In therapy, clients are often embarrassed about disclosing their most painful events, where this emotion binds their representation of an event, their appraisal of the event, and their attendant physiological response. Anticipating talking to a therapist can prompt fear of embarrassment as well as fear of reviving the negative emotions that accompanied the trauma such as Steve's sexual abuse. Joan's nested emotions include feeling guilty about being needy.

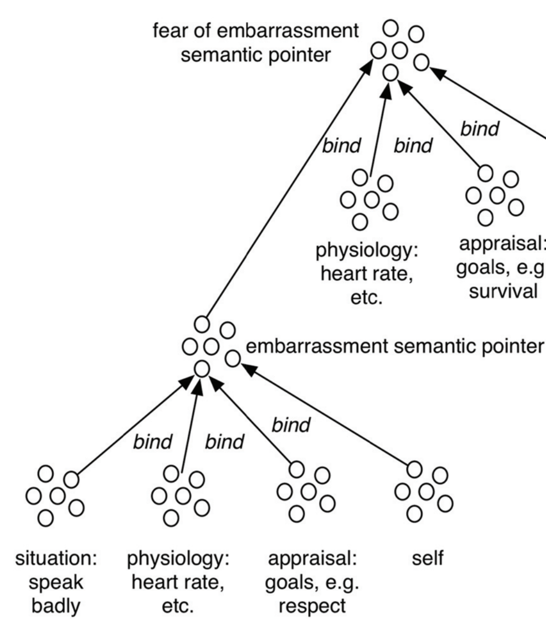
Dispositional Emotions

The processes just described handle the generation of occurrent emotions felt at a particular time. People also have dispositional emotions such as loving family members and fearing snakes; these thoughts are not constantly present but can quickly be generated in particular situations. Because the semantic pointer theory of emotions employs neural mechanisms, it can draw on theories of neural storage to explain dispositional emotions.

What is the connection between short-term emotional change (seconds or hours) and long-term changes (days, months, years)? Anxiety, for example, can be an emotion such as being anxious about a deadline, a diffuse mood lasting a day, or a personality disorder lasting years. Philosophers commonly distinguish between occurrent and dispositional beliefs (Schwitzgebel, 2015), and the same distinction applies to emotions. An occurrent emotion is one happening at a particular time, for example when you are feeling happy. On the other hand, to say that you are a happy person is to say that you have a disposition to have the occurrent emotion of happiness in response to some situations.

Some emotions can be short, over in a few seconds, for example, surprise and disgust; but others can persist for hours, for example, sadness and hatred (Verduyn & Lavrijsen, 2015). Why

Figure 3
Fear of Embarrassment as a Semantic Pointer That Binds Another Semantic Pointer



they end quickly or slowly should also be explainable by a theory of emotional change.

Theories of emotion usually apply to changes in occurrent emotions, but a complete theory should also explain changes in dispositional emotions. If you fall in love with someone, that does not mean that you love them every moment, only that you have a disposition to feel love for them at particular times. Similarly, if psychotherapy helps you to be less depressed and more happy, this change does not mean that you are constantly happy, only that you tend to become happy more often than before. A good theory of emotion should be able to explain both why you become happy at a particular time and also why at times in your life you become more inclined to experience happiness.

In the brain, the relevant mechanisms are the ones used in the formation of memories by the alterations of neural networks. The simplest alteration is the strengthening and weakening of the synapses that connect neurons, which occurs by molecular processes such as changes in the receptors for the neurotransmitter glutamate. Memories are also affected by the generation of new neurons, thousands of which are formed in the hippocampus every day. Another important memory process is the translation of daily memories in the hippocampus into more long-term storage in the cortex. Lane et al. (2015) describe how memory consolidation can contribute to psychotherapeutic change through emotional arousal.

Acquiring the disposition to love someone or fear something amounts to gaining neurons and synapses that enable the brain to generate corresponding semantic pointers. For example, suppose previous experience has led to synaptic connections between neural groups for snakes and for survival goals, as well as for physiological changes in heart rate and cortisol levels. Then presentation of the snake will quickly generate the appraisals and physiological representations that get bound into the semantic pointer that constitutes the intense occurrent emotion of fearing snakes. Thus, the mechanisms of learning, memory, and semantic pointer formation explain how emotions can be both dispositional and occurrent. For example, Steve's bitterness about his abuse is occurrent when he is thinking about it and dispositional when he is thinking about something else but would feel bitter if reminded of the trauma.

Explaining how dispositional emotions change is again based on learning and memory. For example, if you have a series of pleasant experiences with snakes, then the neural networks that generated the current experiences of fear would be changed. Behavioral therapy using desensitization operates in just this way when people are exposed gradually to previously noxious objects and learn to accept rather than fear them. Similarly, the dispositional change of falling in love with somebody means acquiring new memories, neurons, and synaptic connections that are poised to generate neural representations and semantic pointers in response to the object of one's affection.

But what are dispositions? Philosophers often treat dispositions as counterfactual conditionals (Choi & Fara, 2018). For example, to say that a teaspoon of salt has a disposition to dissolve in water is to say that *if* it had been placed in water then it *would* have dissolved. Unfortunately, the standard philosophical way of dealing with such counterfactuals is in terms of possible worlds. To say that the salt would have dissolved is to say that there is some possible world similar to ours in which it does dissolve. This characterization of dispositions in terms of counterfactuals and possible worlds is useless both psychologically and physically. It shows no understanding of the physics of salt and taps into obscure metaphysics about possible worlds rather than into the psychology of how people think about dispositions.

A better way of understanding counterfactuals comes from the artificial intelligence researcher Judah Pearl (2000; Thagard, 2019c). Pearl says that you can evaluate a counterfactual by considering a causal model that shows how different factors interact with each other. For example, the mechanism of solubility of table salt is well understood because we know that sodium chloride results from binding positive sodium atoms with negative chlorine atoms. When salt is placed in water, the ions separate. Knowledge of these mechanisms justifies the conclusion that salt is soluble even if it does not actually get placed in water. We have reason to believe that a counterfactual is true when we know underlying mechanisms that predict the results of various manipulations.

Similarly, we can make inferences about dispositional emotions based on the mechanisms that underlie them, including learning, memory, neural firing, binding, and competition. Moods can also be understood as dispositions based on mechanisms.

Moods

Moods are different from emotions in three main respects. First, moods tend to be much more long-lasting, going for hours or days, whereas an emotion may only last minutes. Second, emotions are about something specific, such as a person or situation, but moods are much more diffuse with no identifiable object. A mood is a general feeling, not a reaction to a particular situation. Third, moods are not as intense as emotions, which can be strong feelings such as exhilaration, terror, or despair. In contrast, you might not be consciously aware that you are in a good or bad mood until you reflect on your response to situations. Steve and Joan both experience depressed mood even when they are not feeling specific emotions about their problems.

Moods and emotions are linked. When you are in a bad mood, you are inclined to have negative emotions such as being sad, angry, or afraid about something. But when you are in a good mood, you are inclined to have positive emotions such as being happy or hopeful about something. The nature of emotions should inform us about the nature of moods. Moods are dispositions to have emotions, based on the same mechanisms.

Being in a mood is having processes going on in your body, and in your brain's unconscious appraisals of situations, which together produce particular kinds of emotions in response to particular kinds of situations. For example, if you are in a good mood, then your physiology and background thinking are operating in ways that incline you to interpret new situations positively. More specifically, if you have some of the physiological hallmarks of good emotions such as stable heart rates, breathing rates, and cortisol levels, and if you have been making mostly positive appraisals about things, then these background processes represented in your brain will make you more likely to respond to new situations with positive emotions. The representation of the new situation and its appraisal will be different but the physiology will influence the new semantic pointer.

Then mood onset is explained in the same way as the beginnings of occurrent emotions, because of changes in neural representations of situations, appraisals, and physiology. Unlike emotions, no semantic pointers corresponding to specific moods are formed about situations, but the background neural activity involving appraisals and physiology inclines the brain to respond to situations with particular position emotions (good mood) or particular negative emotions (bad mood). Shifting from a good mood to a bad mood requires the mechanisms of emotion switching that we now describe.

Emotion Switching

Here are some simple examples of emotion switching. A child is playing happily with a toy, but another child snatches the toy away, provoking a temper tantrum that is a combination of anger and fear. But the child becomes happy again when given another more appealing toy to play with. A romantic couple is having an enjoyable conversation when the topic suddenly shifts to disagreeable matters of discord, such as finances, in-laws, or sex. Then they become distracted by thoughts of a coming joint vacation and the discord that was bordering on anger shifts to affection. A therapist can bring about a switch in a client's occurrent emotions by various means including supportive empathy and sympathy, compliments, and humor.

Sometimes, such shifts are merely a matter of one emotion fading away and another coming into operation, just cessation and onset covered by the accounts already given. But sometimes emotion changes are more sudden when a new emotion arises that replaces a previous one, for example when Joan becomes happy at the therapist's suggestion about new ways of organizing her life which replaces feeling anxious.

The semantic pointer theory of emotions shows how such emotional switches can take place. The major mechanism is competition, with the neurons constituting one semantic pointer inhibiting the neurons in the other semantic pointer, resulting in the new semantic pointer taking over conscious experience. For example, in the child case the anger semantic pointer formed in response to the toy being taken away quickly outcompetes the previously established happiness semantic pointer.

What makes the new semantic pointer stronger than the old one? The firing pattern in the new semantic pointer is formed in the same way as the old one, through bindings of representations of the situation, the appraisal, and the physiology. Changing these neural representations can lead the new emotion to dominate the previous one. In the toddler case, the child no longer has the toy, the child's goals have gone from being satisfied to be thwarted, and the physiology shifts from patterns typical of happiness to those typical of anger. In the couple case, the discussion shifts from amicable to angry to agreeable for the same reasons. In Joan's case, her appraisal and physiology change when she realizes that better organization can help her cope with the demands of work and family.

Different semantic pointers become dominant because of changes in representations of the situation (including the other person), appraisals by each person of the goal-relevance situation, and their physiological states. This shift is not just cessation and onset because of the key role of competition between the new and old semantic pointers. Mixed emotions are ones where the competition is indecisive, so that two or more emotions end up roughly in a tie and therefore are felt together.

The POEM computational model of emotion simulates a simple kind of emotional switching based on reappraisal (Kajić et al., 2019). One simulation models experience of fear when seeing a snake as resulting from the negative linguistic evaluation of snakes. When the simulation adds to the representation of the situation the semantic pointer for glass, it activates the semantic pointer for zoo, which previous experience gives a much more positive evaluation. Then the neural network shifts to an overall interpretation that carries with it more positive emotions such as interest and happiness.

Switching Patterns of Emotions

Often, emotions do not occur in isolation from each other but can be bundled together into patterns. For example, a compatible couple may sometimes share a pattern of emotions that includes love, happiness, and respect, whereas feuding couples may regard each other with anger, hatred, and even contempt. Steve's abuse led to a pattern of connected emotions that included shame, bitterness, and distrust. Joan's problems led to a pattern that included anxiety and depression. Emotional change can therefore require replacing one pattern of emotions with another.

Psychotherapy often needs to treat patterns of emotions when clients seek help for problems such as depression, anxiety, grief, posttraumatic stress disorder, and personality disorders. Table 3 specifies the emotional changes needed for treatment of five common disorders for which people seek help from psychotherapeutic assistance. For each of them, the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (*DSM–5*), lists undesirable emotions that are symptoms of the disorder. The most common disorder is depression, which is characterized by negative emotions such as

Table 3

Transitions in Mental Problems for Which People Seek Therapy in Common Disorders, as Specified in DSM–5 (American Psychiatric Association, 2013)

Disorder	Transition from	Transition to
Major depressive disorder	Sad, empty, hopeless, tearful, worthless, guilty, diminished pleasure	Happy, meaningful, hopeful, pleasure, self-worth
Generalized anxiety disorder	Anxious, worried, restless, irritable	Calm, confident, contented, relieved
Post-traumatic stress disorder	Upset, distressed, self-blame, shame, lack of pleasure, irritable hypervigilant	Calm, accepting, joyful, cheerful, defiant
Persistent complex bereavement disorder	Yearning, longing, sorrow, distress, distrust, depression, loneliness, emptiness	Accepting, cheerful, trusting, happy, social, meaningful
Borderline personality disorder	Fear of abandonment, impulsivity, suicidal behavior, mood instability, emptiness, anger	Confident, calm, happy, stable, meaningful

sadness, emptiness, hopelessness, worthlessness, and guilt. For people to recover from depression, they need to remove or dramatically reduce these unpleasant feelings, and ideally to replace them with their positive opposites such as happiness, meaningfulness, hopefulness, and a sense of self-worth. Thagard and Larocque (2018) provide a model of mental health assessment based on explanatory coherence rather than symptom matching.

Emotions are usually considered as isolated experiences, but the disorders that lead people to seek treatment indicate ways in which bundled emotions can be interconnected. For example, the sadness, hopelessness, worthlessness, and guilt that often accompany depression can be interconnected through all of the neural representations that go into semantic pointers for those emotions. People may feel sad and guilty about the same events, for example the death of a loved one. Hopelessness and worthlessness also concern similar situations concerning a person's place in the world.

In addition to being about the same situations, emotions in a pattern may involve similar appraisals. Sadness, hopelessness, and worthlessness are all based on appraisals that people are not accomplishing their goals. Moreover, these emotions may have similar physiological patterns that go with feeling bad. However, the interconnections among these bundled emotions does not mean that they are all the same, because there may be some variations in the situations, appraisals, and physiology that get bound together into their respective semantic pointers.

Psychotherapists make systematic efforts to change these patterns of emotions. There are many different techniques used for this purpose, including psychodynamic, cognitive–behavioral, and emotion-focused therapies (Greenberg, 2015, 2016). The particular theoretical approach is less important than the ability of a therapist to establish a therapeutic relationship with clients and help them overcome their emotional problems (Castonguay & Hill, 2017).

Therapists cannot operate on semantic pointers directly but can work with the client to change representations of situations, appraisals, and physiology. For example, they can encourage clients to think about aspects of their situation that are not so desperate as what is making them feel sad and hopeless. People may feel sad about their relationship status but nevertheless form alternative emotions based on work success or enjoyable pastimes. Greenberg (2015, 2016) describes ways in which psychotherapists can help people transform negative emotions into positive ones, for example grief into laughter, anxiety into warmth, and shame into pride. His slogan is "emotions change emotions," and coherence mechanisms and semantic pointers explain how.

A major part of how therapists help clients change their emotions is by altering their appraisals of the situation. Appraisals are based in part on beliefs on the extent to which people are capable of accomplishing their goals given current situations. Interactions with therapist can help people to recognize that some of these beliefs about themselves and the world are ill-founded. Revising these beliefs then can lead to different appraisals that then contribute to different patterns of emotions.

More radically, appraisals can be improved by helping people to change their goals. Some goals are simply unrealistic because their accomplishment is beyond the capabilities of the client. Replacing unrealizable goals by ones that are within the range of accomplishment can then help the client to see that some degree of satisfaction can be achieved, for example when Joan seeks a better balance between work and family. Sometimes replacement of unrealistic goals by achievable ones can be furthered by realization that the painful goals were not initiated by the client, but rather by a parent, spouse, or other authority figure who shaped or controlled the client's agenda.

Finally, psychotherapists can help clients to improve their physiological states. Meditation and deep breathing can have calming effects, and increased exercise can change bodily markers such as heart rate, breathing rate, blood sugar, and cortisol levels. Antidepressants can complement therapy by changing levels of important neurotransmitters such as serotonin, dopamine, and norepinephrine, and also by bringing about other neurochemical changes such as increasing BDNF (brain-derived neurotrophic factor) and encouraging neurogenesis.

Changes in patterns of emotions that occur in psychotherapy can be both occurrent and dispositional. In a single session, a psychotherapist can help people to feel better about themselves and therefore carry out short-term emotional switches from feeling sad and worthless to feeling hopeful about improving their mental state. But the goal of psychotherapy is long-term change, which requires modifying dispositions so that people are more inclined to be positive about their lives rather than negative.

These dispositional changes require the same neural mechanisms described above for single emotions, leading to different packages of neurons and synapses that generate patterns of firing corresponding to different semantic pointers. Psychologically, these permanent changes require very different configurations of neurons and connections about them, corresponding to different views of the world and the self.

Psychotherapy is just one of the ways in which people can replace patterns of emotions by others. People also sometimes manage to transform their outlooks on life through experiences such as self-help, religion, political engagement, and other activities that dramatically alter their beliefs, goals, and physiological states. In all of these cases, people end up with different patterns of emotions as the result of occurrent and dispositional changes that are explained by the neural mechanisms that generate semantic pointers.

New Simulations

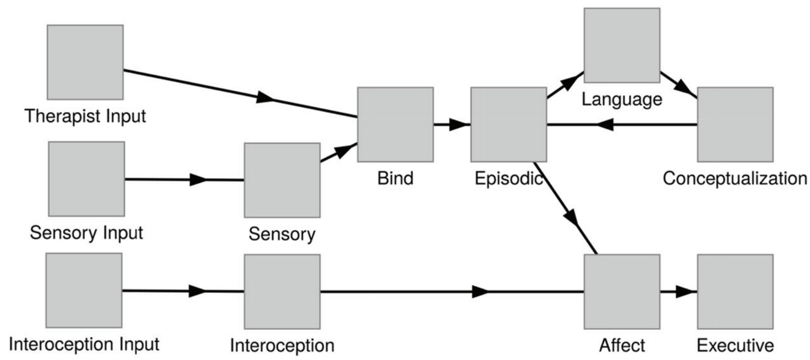
To model emotional change in psychotherapy, we have used the neural engineering simulator Nengo for new simulations of the Steve and Joan cases, the latter described in the online supplemental material. Figure 4 shows the structure of the POEM model used to simulate the impact of a therapist on the emotions and actions of a client. The verbal and nonverbal behavior of the therapist is represented by a vector that is translated into a pattern of firing in a group of spiking neurons. This pattern is bound by convolution with other sensory input that gets combined with interoception input from the body's internal sensors for heart rate and other physiological changes. Language and concepts also contribute to interpretation of the current episode to change affect by introducing specific emotions that lead to actions as the result of executive decisions.

For the Steve case, his initial mental state is modeled as the semantic pointer:

$$\mathrm{VICTIM} \circledast \mathrm{C1} + \mathrm{BITTER} \circledast \mathrm{C2} + \mathrm{HELPLESS} \circledast \mathrm{C3}$$

Here "$\circledast$" is the circular convolution operator that binds two vectors such as ones for VICTIM and for C1. The interpretation of

Figure 4
POEM Model Used to Simulate the Steve and Joan Cases



Note. Adapted from "The Semantic Pointer Theory of Emotions," by I. Kajić, T. Schröder, T. C. Stewart, and P. Thagard, 2019, *Cognitive Systems Research, 58,* p. 42 (https://doi.org/10.1016/j.cogsys.2019.04.007). Copyright 2019 by Elsevier. Reprinted with permission.

what each individual vector represents depends on the specific modeling context, such as when syntactic markers indicate parts of speech such as Subject or Object, or when positions in a serial memory task indicate each word as bound to as specific position in a list. Here, vectors are used to indicate abstract concepts that people associate with themselves. Binding each individual concept with a marker (C1, C2, C3) gives a general mechanism that can replace a semantic pointer with a different one, which would not be possible if we used simple vector addition such as VICTIM+BITTER+HELPLESS.

The "+" operator adds the vectors together to produce a vector for Steve as "bitter, helpless victim" that translates into a neural pattern that generates negative emotion. This input to the sensory network can be interpreted as thoughts elicited by a situation that prompts Steve to think of himself as victim, helpless, and bitter. We do not model specific sensory information that elicits those thoughts but rather the resulting thoughts that are empirically acquired by the patient. The interoception neural input is initially presented with a vector, FELT_HEARTBEAT_GETTING_FASTER, which simulates the physiological response related to the processing of current information from the body.

The therapist helps Steve by guiding him to see himself as intelligent, handsome, and hardworking rather than as a helpless victim. In the simulation, the therapist input changes the neural representation to:

$$\text{VICTIM} \circledast (\sim\text{HANDSOME}) \circledast \text{C1} + \text{BITTER} \circledast (\sim\text{INTELLIGENT}) \circledast \text{C2} + \text{HELPLESS} \circledast (\sim\text{HARDWORKING}) \circledast \text{C3}.$$

This input replaces the semantic pointer for VICTIM with HANDSOME, BITTER with INTELLIGENT, and HELPLESS with HARDWORKING when its inverse is bound with the semantic pointer represented in the sensory network. The resulting semantic pointer, HANDSOME + INTELLIGENT + HARDWORKING, outcompetes the previous semantic pointer for VICTIM + BITTER + HELPLESS. This change in representation prompts a shift to positive emotions.

This simulation does not model the complex dynamics of psychotherapy, but it proposes computational ingredients that underlie emotional change. For example, by modeling the notion of "self' as a composition of semantic pointers, we assume (a) that individuals are able to identify the source of their emotional state with the help of a psychotherapist, (b) that attributes associated with that state depend on each other, and (c) that the therapist, through the interaction, has the ability to help the individual to replace those negative views. A prediction that follows from this formulation is that to replace negative emotions, it helps first to identify them rather than just to offer alternative, positive views of self.

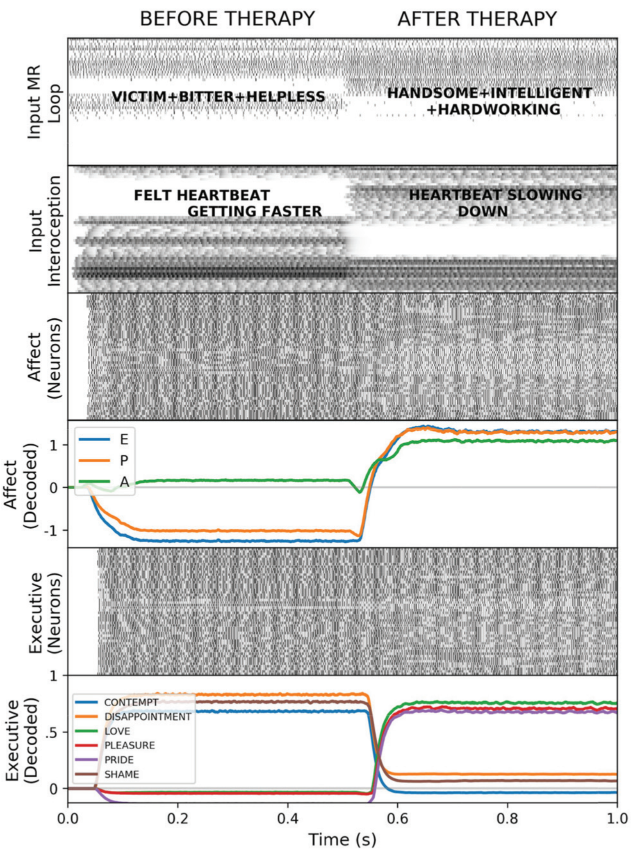
Figure 5 shows a subset of neural activity as spikes emitted by individual neurons and vectors decoded from all spikes. The figure shows how the change in Therapist Input and Interoception Input led to output in the executive network that changes from *disappointment*, *shame*, and *contempt* (negative emotions) to *love*, *pleasure*, and *pride* (positive emotions). The first panel shows the inputs to the memory-reasoning (MR) loop (consisting of Episodic, Language, and Conceptualization networks, as shown in Figure 4), which are the result of binding of semantic pointers represented in the sensory and therapist inputs. The second panel shows neural activity representing the interoception input. The third panel shows the activity of neurons in the affect network and the similarity between the vector represented by that neural population with the ideal vectors for evaluation (E), potency (P), and arousal (A). The activity for the executive network shows that emotions with negative valence are replaced by emotions with positive valence following the change in the therapist input. Because of length limitations, mathematical details and description of the Joan model are relegated to Supplement 2 of the online supplemental materials. A fuller account of the underlying mathematics can be found in Kajić et al. (2019). Thagard and Larocque (2020) contains more details about the conceptual structure of the Steve and Joan cases.

Discussion

We have shown that the semantic pointer theory of emotions explains a wide range of kinds of emotional change. This theory

Figure 5

Neural Activity and Similarities of Vectors Decoded From Neurons in Different Parts of the POEM Model Simulating Emotional Change in Steve



Note. Only a subset of spiking activity over time from 50 neurons in each of the networks is shown, and the labels are for vectors represented by neural activity. See the online article for the color version of this figure.

builds on the insights of appraisal, physiological, and cultural theories of emotion but synthesizes them using neural mechanisms that operate in many mental processes. For emotion, mechanisms produce the formation of patterns of firing in groups of neurons that are capable of representing aspects of selves, situations, appraisals, and physiological changes. These patterns of neural firing are bound into more complex patterns called semantic pointers which capture a broad range of emotions and compete to represent the overall current state of a person.

This theory covers both simple emotional states and more complex ones that are mixed and nested. It applies both to occurrent emotions presently happening and to dispositions to have emotions over longer periods of time. The theory covers changes in individual emotions but also changes in patterns of emotions such as those that need to be dealt with in psychotherapy.

The semantic pointer theory of emotion and emotional change is sufficiently precise that it can be stated mathematically and modeled computationally. The POEM model provides an approximation of the changes that occur in real-life psychotherapy such as the Steve and Joan cases, using the same mechanisms to produce two new simulations in addition to the six presented previously. The result is not just a description of emotional change but an explanation of how it occurs.

Nevertheless, many of the phenomena of emotional change described in this paper have yet to be simulated. Supplement 3 in the online supplemental materials lists a range of phenomena that should be open to simulation. We also need to say more about the cultural and social aspects of emotions. It should be possible to spell out in much more detail how the interactions with other people such as friends, families, and psychotherapists help to induce emotional change. For this purpose, the understanding of verbal and nonverbal communication as semantic pointer transfer is useful (Thagard, 2019b).

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (Eds.). (2016). *Handbook of emotions* (4th ed.). Guilford Press.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Routledge.

Blouw, P., Solodkin, E., Thagard, P., & Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive Science*, *40*(5), 1128–1162. https://doi.org/10.1111/cogs.12265

Castonguay, L. G., & Hill, C. E. (Eds.). (2017). *How and why are some therapists better than others? Understanding therapist effects*. American Psychological Association.

Choi, S., & Fara, M. (2018). *Dispositions*. https://plato.stanford.edu/entries/dispositions/

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

Crawford, E., Gingerich, M., & Eliasmith, C. (2016). Biologically plausible, human-scale knowledge representation. *Cognitive Science*, *40*(4), 782–821. https://doi.org/10.1111/cogs.12261

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Henry Holt.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. MIT Press.

Greenberg, L. S. (2015). *Emotion-focused therapy: Coaching clients to work through their feelings*. American Psychological Association.

Greenberg, L. S. (2016). The clinical application of emotion in psychotherapy. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 670–684). Guilford Press.

Harré, R. (Ed.). (1989). *The social construction of emotions*. Blackwell.

Heavey, C. L., Lefforge, N. L., Lapping-Carr, L., & Hurlburt, R. T. (2017). Mixed emotions: Toward a phenomenology of blended and multiple feelings. *Emotion Review*, *9*(2), 105–110. https://doi.org/10.1177/1754073916639661

Heise, D. R. (2007). *Expressive order: Confirming sentiments in social actions*. Springer.

Hudlicka, E. (2011). Guidelines for designing computational models of emotions. *International Journal of Synthetic Emotions*, *2*(1), 26–79.

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188–205.

Kajić, I., Schröder, T., Stewart, T. C., & Thagard, P. (2019). The semantic pointer theory of emotions. *Cognitive Systems Research*, *58*, 35–53. https://doi.org/10.1016/j.cogsys.2019.04.007

Keltner, D., Oatley, K., & Jenkins, J. M. (2018). *Understanding emotions* (4th ed.). Wiley.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, *38*, 1–80. https://doi.org/10.1017/S0140525X14000041

Marsella, S., & Gratch, J. (2014). Computationally modeling human emotion. *Communications of the ACM*, *57*(12), 56–67. https://doi.org/10.1145/2631912

Marsella, S., & Gratch, J. (2016). Computational models of emotion as psychological tools. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 113–129). Guilford Press.

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Nussbaum, M. (2001). *Upheavals of thought*. Cambridge University Press.

Oatley, K. (1992). *Best laid schemes: The psychology of emotions*. Cambridge University Press.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Plate, T. (2003). *Holographic reduced representations*. CSLI.

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Reisenzein, R. (2017). Varieties of cognition-arousal theory. *Emotion Review*, *9*(1), 17–26. https://doi.org/10.1177/1754073916639665

Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., & Meyer, J.-J. C. (2013). Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange. *IEEE Transactions on Affective Computing*, *4*(3), 246–266. https://doi.org/10.1109/T-AFFC.2013.14

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion*. Oxford University Press.

Schwitzgebel, E. (2015). *Belief*. https://plato.stanford.edu/entries/belief/

Thagard, P. (2012). *The cognitive science of science: Explanation, discovery, and conceptual change*. MIT Press.

Thagard, P. (2019a). *Brain-mind: From neurons to consciousness and creativity*. Oxford University Press.

Thagard, P. (2019b). *Mind-society: From brains to social sciences and professions*. Oxford University Press.

Thagard, P. (2019c). *Natural philosophy: From social brains to knowledge, reality, morality, and beauty*. Oxford University Press.

Thagard, P., & Aubie, B. (2008). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition*, *17*(3), 811–834. https://doi.org/10.1016/j.concog.2007.05.014

Thagard, P., & Larocque, L. (2018). Mental health assessment: Inference, explanation, and coherence. *Journal of Evaluation in Clinical Practice*, *24*(3), 649–654. https://doi.org/10.1111/jep.12885

Thagard, P., & Larocque, L. (2020). *Mapping emotional change in psychotherapy*. https://paulthagard.files.wordpress.com/2020/07/thagard.map-emo.20jl.pdf

Thagard, P., & Schröder, T. (2014). Emotions as semantic pointers: Constructive neural mechanisms. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotions* (pp. 144–167). Guilford Press.

Thagard, P., & Stewart, T. C. (2014). Two theories of consciousness: Semantic pointer competition vs. information integration. *Consciousness and Cognition*, *30*, 73–90. https://doi.org/10.1016/j.concog.2014.07.001

Verduyn, P., & Lavrijsen, S. (2015). Which emotions last longest and why: The role of event importance and rumination. *Motivation and Emotion*, *39*(1), 119–127. https://doi.org/10.1007/s11031-014-9445-y

Wagar, B. M., & Thagard, P. (2004). Spiking Phineas Gage: A neurocomputational theory of cognitive-affective integration in decision making. *Psychological Review*, *111*(1), 67–79. https://doi.org/10.1037/0033-295X.111.1.67

Received June 17, 2020
Revision received January 28, 2021
Accepted February 3, 2021 ■