CHAPTER 6

Discrete emotions

From folk psychology to causal mechanisms

Andrea Scarantino
Department of Philosophy, Neuroscience Institute, Georgia State University

1. Introduction

Panksepp and Russell are eager to find common ground in their fascinating debate. To combine the theoretical insights each approach has to offer, one central issue of contention must be resolved: Are discrete emotions genuine causal mechanisms with respect to the components commonly associated with them? By "discrete emotion", I will mean any emotion episode designated by an individually separate and distinct category (e.g. fear, anger, sadness). With respect to fear, for instance, we can ask: Does it bring about the heart rate changes, action tendencies, facial expressions and feelings commonly associated with it?

The way we settle the issue of emotional causation has key methodological implications. If discrete emotions are not genuine causal mechanisms, as posited by Russell, they should not be allowed into affective science's basic explanatory toolbox. If at least some discrete emotions are genuine causal mechanisms, as suggested by Panksepp instead, providing emotion-based explanations of physiological, behavioral, expressive, and phenomenological events should be one of affective science's primary tasks.

In my commentary, I side with Panksepp on the issue of emotional causation. But I will argue that there are key lessons to be learned from Russell's opposition to standard accounts of discrete emotions. In the spirit of fostering discussion, I will conclude by proposing a revision of Panksepp's discrete model that takes into account Russell's critique.

To avoid mixing substantive disagreements with disagreements that are merely terminological, I begin with a quick summary of the two accounts, presented in as theory-neutral a fashion as possible.

2. Panksepp on primary, secondary and tertiary affects

According to Panksepp, "affects are feelings that guide our thoughts and our actions" (32).[1] On this view, affects are by definition associated with a subjective experience, i.e. a feeling. In other places, however, Panksepp uses the term "affects" more broadly to designate emotional operating systems, on the assumption that such systems produce feelings once activated.

To increase clarity, I distinguish between the *functional notion* of an "affective system", defined by its ability to coordinate organismic resources towards an end (e.g. a danger-avoiding system), and the *phenomenological notion* of an "affective feeling", defined by the subjective experience attached to it (e.g. a feeling of fear). Whether affective systems *always* produce affective feelings once activated is a question I will leave open.

Panksepp distinguishes between three major classes of affective feelings: *emotional affective feelings* (e.g. the feeling of fear), *sensorially triggered affective feelings* (e.g. the feeling of pleasure produced by tasting truffles), and *homeostatic affective feelings* (e.g. the feeling of hunger).

Within the class of *emotional affective feelings*, he distinguishes between *primary-process, secondary-process* and *tertiary process* affective feelings. Primary-process affective feelings, a.k.a. "core affects", are produced by "primary-process emotional systems" (33). There are "at least seven" of those, labeled by Panksepp SEEKING, RAGE, FEAR, LUST, CARE, GRIEF/PANIC, and PLAY.

Each primary system is associated with a genetically based subcortical brain network unconditionally triggered by a limited number of sensory/perceptual inputs connected to major life-challenging circumstances. But it can also be conditionally triggered by neutral stimuli. For instance, primary-process systems can combine with general-purpose learning processes such as "classical and instrumental conditioning" (32) to give rise to secondary-process affective feelings. They can also combine with higher cognitive processes such as "perceiving, thinking, ruminating, fantasizing, etc." (32) to give rise to tertiary-process affective feelings.

Panksepp's belief in the existence of primary affective systems is largely due to "our ability to artificially activate various kinds of emotional patterns by applying the appropriate kinds of chemical or electrical stimulation to specific subcortical regions of the brain" (38). For instance, Panskepp argues that the electrical stimulation of the FEAR network can lead animals to manifest fearlike behaviors, from freezing responses to flight responses depending on the intensity of the electric current.

1. Unless otherwise stated, all page numbers refer associated with Panskepp's quotes refer to Chapter 2 in this volume.

Panksepp declares that his research "is almost exclusively devoted to the *primary-process* nature of emotions" (33), in particular to the understanding of their "causal infrastructure" at the neural level.

3. Russell on core affect, affective quality and meta-experience

Russell begins by stating that he will use the term affect "in a narrow sense limited to private subjective conscious feelings" (85).[2] Later on, however, he allows for the possibility that some varieties of affect (e.g. Core Affect) may occur without being attended to, a circumstance that gives rise to unconscious affect. Russell emphasizes that "*affect* must be distinguished from what I call an *emotional episode*" (85), namely a particular event referred to by everyday English words such as "emotion", or "fear", or "anger".

Consider Sally's fear of a bear in the woods last Saturday at 10 am. For Russell, this episode includes an "affect", i.e. an affective feeling, jointly with "a sequence of other components as well, such as Sally's perception of and appraisal of the bear, all the peripheral physiological changes that occurred, facial and vocal changes, her behavioral reaction of freezing and then fleeing, and all the neural processes that underlie these changes" (85). As we shall see, Russell does not think that there is an internal emotion mechanism responsible for the co-instantiation of such components.

Russell distinguishes "three different types of affect, without claiming that these three exhaust the domain: Core Affect, Affective Quality, and Emotional Meta-experience" (85). Core affect is "a neurophysiological state that is consciously accessible as a simple, non-reflective feeling that is an integral blend of hedonic (pleasure–displeasure) and arousal (sleepy–activated) values" (Russell, 2003: 147).

Like Panksepp's primary-process emotional systems, Core Affect is assumed to be associated with specific brain networks, and Russell makes some suggestions as to what such networks may be (cf. Pfaff, 2006; Smith et al., 2010). Core Affect need not be directed at anything, but it is "an elemental building block of other psychological events" (88) directed at something, including discrete emotions. Sally's fear episode, for instance, will include as its proper part a feeling of high arousal and high displeasure directed at the bear.

The primary, but by no means exclusive, source of evidence for the existence of Core Affect comes from psychometric studies of self-reported moods and

2. Unless otherwise stated, all page numbers associated with Russell's quotes refer to Chapter 3 in this volume..

emotions. These studies have shown that self-reports of how one feels or felt at a particular moment or expects to feel in the future correlate in ways that can be parsimoniously explained along the two dimensions of *valence* and *arousal*.

In a nutshell, subjects tend to report experiencing moods and emotions with the same valence together and moods and emotions with the same arousal together. If they report feeling afraid, for instance, they also tend to report feeling angry. What underlies this surprising correlation, it is proposed, is that feelings of fear and anger typically have in common high displeasure and high arousal.

The second type of affect Russell considers is the "affective quality" of a stimulus, namely "its capacity to change core affect" (Russell, 2003: 149). As with every capacity, the capacity to change Core Affect need not be manifested in order to be possessed. A given sunset on Daytona beach may have had the affective quality of being *soothing* even though no one observed it, or even though the only person observing it was too depressed to be soothed. The sunset is still soothing because it would lead an observer in suitable circumstances to be soothed by it.

The third type of affect distinguished by Russell is "emotional meta-experience", which amounts to the categorization of one's discrete emotion under a concept. "On my account of emotional meta-experience", Russell suggests, "to perceive oneself as afraid is to categorize oneself by means of the concept of fear" (91). Since concepts have correctness conditions for their application, one's meta-experience may be mistaken. Sally may categorize herself as being afraid when she is angry instead. Furthermore, anyone who lacks the concept of fear can't have an emotional meta-experience of fear.

4. Are discrete emotions causal mechanisms?

With these terminological preliminaries out of the way, let us consider what I take to be the central issue of contention between Russell and Panksepp. Whereas Panksepp posits seven "core" affective systems, i.e. SEEKING, RAGE, FEAR, LUST, CARE, GRIEF/PANIC, and PLAY, each associated with a discrete affective feeling, Russell argues that there is only one "core" affective system, i.e. Core Affect, associated with affective feelings resulting from blends of pleasure and arousal values. At first blush, the two accounts are compatible: Panksepp's seven affective systems could in principle co-exist with Russell's Core Affect system. Which of such systems should be called "core" strikes me as being more a matter of terminological predilection than theoretical substance.

But the devil is in the details. Panksepp believes that there are discrete emotions that *cause* the suite of physiological, behavioral, expressive, and phenomenological components associated with them. His seven "core" affective systems

provide an example of causally powerful discrete emotions. Such systems have the "ability to coordinate sets of behavioral and autonomic outputs [and] have a coherent, behavioral-affective coherence promoting infrastructure" (43).

Russell's position is instead that the suite of behavioral, physiological, expressive, and phenomenological components characteristic of episodes of discrete emotion is simply the *evidence* on the basis of which *emotion concepts* are ascribed. His view is that "[w]e need no extra mechanism" that is the emotion to explain why physiological, behavioral, expressive, and phenomenological components co-occur to the (limited) extent that they do.

An important corollary is that *emotion, fear, anger, happiness, sadness* and so on are of scientific interest to Russell merely as *folk concepts*. We can usefully explore which patterns of components are categorized under one folk emotion concept rather than another. This amounts to engaging in a *descriptive project*, whose objective is the "analysis of what [emotion] words mean in everyday discourse" (80). But when we shift to the *prescriptive project*, i.e. the "scientific study of the events to which words such as *emotion, affect,* and *jealousy* refer" (80), folk emotion concepts are no longer suitable. This is why affective science must abandon "traditional ways of thinking" (79) about emotion, and shift to the study of Core Affect, affective quality and emotional meta-experience.

At this juncture we encounter a puzzle. Russell's rejection of the causal import of discrete emotions is grounded in skepticism about the scientific suitability of folk emotion concepts. Everyday emotion concepts, he writes, "have become a source of difficulty in furthering a scientific analysis of emotion" (80). Yet, Panksepp seems to share the same skepticism: "I have pointedly chosen not to use vernacular terms for primary-process emotional systems and their affects" (33).

How did skepticism about folk emotion concepts lead Russell and Panksepp in such different directions? Can their differences be reconciled?

5. What's wrong with folk emotion concepts?

Russell and Panksepp are both convinced that folk emotion concepts are unsuitable for scientific purposes. This puts them in contrast with the great majority of emotion theorists, who characterize their objects of investigation in folk psychological terms. Competing scientific theories of emotion/fear/anger/etc. disagree on content, but they share the assumption that there is nothing wrong with emotion, or fear, or anger as scientific concepts.

Russell begs to disagree. On his view, folk emotion concepts have five fundamental flaws: they "lack defining properties", they admit "borderline cases", they are "culture-specific", they carry "hidden assumptions" from our religious,

philosophical, and intellectual traditions, and they are "too heterogeneous". According to Panksepp, folk emotion concepts have two fundamental flaws: they have "variable meanings due to our socio-cultural diversity" and they generate "part-whole confusions".

I will consider these alleged problems one at a time, and conclude that heterogeneity is the only fatal flaw of folk emotion concepts. The rest of my paper considers what methodological consequences follow from the heterogeneity of folk emotion concepts, specifically with respect to the investigation of discrete causal mechanisms.

5.1 No defining conditions, borderline cases and homeostatic property cluster kinds

Russell rightly emphasizes that it is very difficult to formulate classical definitions of folk emotion concepts such as *emotion* or *fear*, namely definitions consisting of individually necessary and jointly sufficient properties for being an episode of emotion or an episode of fear. Despite centuries of attempts, no theory has been able to unveil agreed upon defining properties.

Furthermore, there seem to be borderline cases of folk emotion concepts, understood as cases that produce widespread disagreement about membership in the language community. For instance, when asked whether calmness is an emotion, 48% of subjects state that it is, and 52% of subjects state that it is not (Fehr and Russell, 1984). This is the sort of response distribution found when we ask whether a 39-year old person is middle-aged. As there does not seem to be a fact of the matter as to whether a 39-year old person is middle aged, there does not seem to be a fact of the matter as to whether or not calmness is an emotion in the folk sense.

The difficulty to find classical definitions and the existence of borderline cases must be explained by a satisfactory theory of folk emotion concepts. Russell's central proposal is that folk emotion concepts are mental representations of scripts, which stand to events as prototypes stand to objects (Fehr and Russell, 1984).

An emotion script "contains prototypical causes, beliefs, physiological reactions, feelings, facial expressions, actions, and consequences" (Fehr and Russell 1984: 482) associated with specific folk emotion categories. For example, the fear script may contain a danger of some kind as a prototypical cause, heightened heartbeats and muscle tension, an unpleasant feeling, a facial expression that includes raised upper eyelids and dropped open jaw, an action tendency of avoidance, and a general physiological preparation for escape. This mental script can be

the basis for the following categorization rule: something is "fear" in the folk sense just in case it achieves enough similarity with the fear script.

This proposal has several virtues. First, it explains why classical definitions are so hard to come by for folk emotion concepts. It would be purely coincidental if all items that achieve enough similarity with some folk emotion E's script happened to share properties that are individually necessary and jointly sufficient for being E. Second, it explains why there are borderline cases of folk emotion categories, which emerge when the similarity threshold to the relevant script is approached but not exceeded.

My point is that lacking a classical definition and having vague boundaries are not obstacles to the scientific suitability of a concept. Biological species concepts are a case in point. There is a great deal of variation among members of the same species. Certain characters are marked as 'highly variable'; some species are divided into several geographic 'races'; some larger groups of birds vary in such complex and overlapping ways that they are denoted a 'species complex'. All this was well known to Darwin and it is what evolutionary theory should lead us to expect.

Yet, despite not being classically definable and having blurred edges, species concepts are scientifically suitable. A good way to make sense of this fact is to consider species concepts to be natural homeostatic property cluster (HPC) kinds (Boyd, 1999). Boyd introduced this influential theory of scientific kinds to account for the fact that many of the central theoretical constructs in the special sciences have internal variability and lack sharp boundaries. According to Boyd, "[t]he natural definition of…homeostatic property clusters kinds is determined by the members of a cluster of often co-occurring properties and by the ("homeostatic") mechanisms that bring about their co-occurrence" (Boyd, 1999: 141).

On the HPC view, something is an instance of a biological species concept like C*anis Lupus* by virtue of sharing a cluster of "often co-occurring" properties, but there are no properties individually necessary and jointly sufficient for membership. Furthermore, some individuals may have just enough of the co-occurring properties to place them on the borderline for membership. These co-occurring properties tend to cluster together by virtue of causal mechanisms such as interbreeding, shared ancestors and exposure to common selection pressures.

On account of such "homeostatic mechanisms", the members of the *Canis Lupus* species tend to imperfectly share properties. Finally, the often co-occurring properties of members of this species are of central interest to biology, in the sense that biologists can formulate lots of reliable inductions and explanations about Canis Lupus.

This example demonstrates that we can find scientifically suitable categories that are *not* classically definable and admit borderline cases. The converse holds as well. We can find scientifically unsuitable categories that *are* classically definable and lack borderline cases. The ancient astronomical concept of a *superlunary object* has classical defining conditions – an object is superlunary if and only if it is located beyond the orbit of the moon – and it lacks blurred edges, but it is clearly unsuitable for contemporary astronomy.

A critic may retort that neither biological species concepts nor the concept of a superlunary object are vernacular concepts, unlike folk emotion concepts. This is true, but irrelevant to the point being made, which is that lacking defining conditions and having borderline cases are not *per se* obstacles to the scientific suitability of folk emotion categories. If they were, we could not find scientifically suitable concepts, whether grounded in the vernacular or not, with such properties. Since we can find them, the case against folk emotion concepts must find support elsewhere.

5.2 Cultural specificity and hidden assumptions

Russell and Panksepp have both suggested that there is another source of trouble with folk emotion concepts, namely that they are "culture-specific" (Russell, 80) or have "variable meanings due to socio-cultural diversity" (Panksepp, 33). It is certainly true that emotion concepts differ to some extent across cultures.

As Russell reminds us by way of example, Gidjingali speakers do not have any concept that directly translates with "fear", combining instead shame and fear into the culture-specific concept of "gurakadj" (110). Languages also differ in the overall richness of their emotion vocabulary, and in the extent to which specific emotions are hypercognized or hypocognized, namely associated with a lexically rich or poor conceptual framework (Levy, 1973).

These sorts of cultural differences are relevant with respect to affective phenomena that depend on concept use. An example is Russell 's notion of the *meta-experience* of an emotion, understood as the categorization of one's emotion under concepts. A Gidjingali speaker who lacks the concept of fear would not be able to categorize herself under the *fear* concept (on the other hand, she would be able to categorize herself under the *gurakadj* concept).

This being said, I do not think that cultural differences have any impact on the scientific suitability of folk emotion concepts. An analogy may help. Consider the category we call in English "gold". Chemists consider it scientifically suitable for chemistry, and they do so because items in the extension of the "gold" category – the extension of a category C is the set of things that fall into C – share lots of

properties relevant to induction and explanation in chemistry. By contrast, items in the extension of the "jade" category do not, because the category comprises two very different minerals called "jadeite" and "nephrite" which just happen to look alike, and are labeled "jade" simply on account of how they look.

These judgments about the scientific suitability of "gold" and "jade" depend entirely on the nature of their extensions *in English*. The point is that the extension of a category in a given language is not affected by whether *other languages* have or lack a category inter-translatable with it. Suppose that a speaker of Gidjingali were to describe gold as "gilder", a concept that includes what in English we call "gold" and "silver". Quite clearly, this would not make "gold" a scientifically unsuitable category in English. Conversely, suppose that a speaker of Gidjingali were to lack the "jade" category, and distinguished in all cases between "jadeite" and "nephrite". Once again, this would not make what we call "jade" in English a scientifically suitable category.

Russell also remarked that folk emotion concepts bring with them "the assumptions of our religious, philosophical, and intellectual traditions, including such dubious distinctions as that between body and mind or that between reason and emotion" (80). Let us assume for the sake of argument that these are in fact misleading assumptions about emotions.

They would have to be accounted for by what Russell calls the *descriptive project*, which aims to reconstruct the "meaning" of folk emotion terms. But these mistaken assumptions do not affect the *prescriptive project*, which focuses on the scientific study of the events to which the terms refer. As the mistaken popular assumption that if something is water it must be liquid does not stand in the way of the scientific study of water, so the (possibly) mistaken popular assumption that if something is an emotion it must be irrational does not stand in the way of the scientific study of what we call "emotion".

5.3 Part-whole confusions reinterpreted

Panksepp uses capitalized terms to designate his seven primary systems: SEEKING, RAGE, FEAR, LUST, CARE, GRIEF/PANIC, and PLAY. These terms preserve orthographic identity with ordinary lower-case English terms such as "seeking", "rage", "fear" and so on. What is the capitalization supposed to indicate? Panksepp's main rationale for using capitalized letters is "to avoid part-whole confusions" (33). Capitalized letters, he says, only refer "to specifiable brain networks that are important *parts* of intrinsically ambiguous conceptual *wholes*" (33).

This statement is naturally interpreted as follows: the upper-case terms SEEKING, RAGE, FEAR and so on refer to brain networks that are parts of the wholes

designated by the lower-case terms "seeking", "rage", "fear", etc. This would not be a satisfactory rationale for capitalization on two levels.

Firstly, if the aim of the capitalized letters were to refer *merely* to a specifiable brain network, Panksepp's terminological choice would be infelicitous. Language affords the much clearer expressions "SEEKING network", "RAGE network", "FEAR network", and so on.

Secondly, it is not the case that Panksepp's brain networks are always parts of the wholes designated by the lower-case terms "seeking", "rage", "fear". As pointed out by Russell, there are instances of activation of, say, the SEEKING network that are not parts of what we call "seeking" in English. Similarly, there are instances of what we call "seeking" in English that do not involve the SEEKING network as a part.

I propose another interpretation, according to which SEEKING, RAGE, FEAR, LUST, CARE, GRIEF/PANIC, and PLAY refer to discrete *affective systems* that have dedicated brain networks as parts. On this view, Panksepp's networks are not parts of the "intrinsically ambiguous conceptual wholes" designated by lower-case terms, but rather parts of the seven theoretically motivated coordination systems introduced by Panksepp.

Such systems are not folk psychologically defined, but rather defined by a number of characteristics of the brain networks supposed to underlie them (Panksepp 42–44, with my wordings):

1. Primary affective systems are genetically based and their activities are unconditionally triggered by a limited number of sensory/perceptual inputs connected to major life-challenging circumstances
2. Primary affective systems coordinate sets of behavioral and autonomic outputs, including hormonal and immunological parameters, in ways that have proven adaptive in the evolutionary history of the species
3. Primary affective systems gate and modulate incoming sensory inputs relevant to the behavioral and autonomic outputs coordinated
4. The activities of primary affective systems outlast their precipitating circumstances
5. The activities of primary affective systems can be conditionally triggered by neutral stimuli through learning and higher thought
6. The activities of primary affective systems influence, and are influenced by, the activities of the brain networks underlying higher decision making and consciousness
7. Primary affective systems produce affective feelings

Panksepp adds that the activities of primary systems "can occur without the last two attributes" (43), which suggests that (6) and (7) are not strictly necessary for

an affective system to be activated. In light of this caveat, I propose we understand Panksepp's primary systems as homeostatic property cluster kinds (Section 5.1). On this view, properties (1)–(7) constitute a family of imperfectly co-occurring properties brought about by a subcortical brain mechanism.

The problem is that Panksepp did not provide a principled justification for the orthographic identity between the names of his seven primary affective systems and ordinary folk emotion concepts. Why should a "general-purpose appetitive motivational system", which aims to motivate "animals to find and eagerly anticipate all kinds of resources they need for survival" (34–35) be called SEEKING? Why should a system "aroused by frustration arising from any attempts to curtail an animal's freedom of action" and "invigorat[ing] aggressive behaviors" (35) be called RAGE? Why should a system "designed during brain evolution to help animals reduce pain and the possibility of destruction" (36) and leading to fleeing or freezing behaviors be called FEAR? Absent a convincing justification, a better strategy would be to use neologisms for the designation of primary affective systems (e.g. the WIS4 affective system, or the threat-coping system).

I propose two conditions as individually sufficient to justify orthographic identity. The first is what I call the *subset condition*. It is satisfied when the capitalized terms apply to a subset of the episodes named by lower-case terms, and nothing else. FEAR would satisfy the subset condition if all instances of FEAR were instances of folk psychological "fear". In this case, Panksepp's designation would be legitimate, because FEAR would describe what a particular kind of "fear" is like. The capitalization would simultaneously warn the reader that, even though all instances of FEAR are instances of "fear", the converse does not hold: there are instances of "fear" that are *not* instances of FEAR. Russell provides one such example – child's separation anxiety.

The second is what I call the *significant overlap condition*. It is satisfied when the capitalized terms apply to significantly more episodes named by lower-case letters than they apply to episodes to which the lower-case letters do not apply. The significant overlap condition would be fulfilled by FEAR even if there were instances of FEAR that are not instances of "fear", as long as significantly more instances of "fear" than instances of non-"fear" were included in FEAR. Panksepp's designation would still strike me as legitimate in this case. This is because *most* FEAR episodes, rather than *all* FEAR episodes as in the subset condition, would still qualify as "fear" episodes.

Admittedly, this second condition is less clear-cut than the first. It relies on a vague notion such as including "significantly more" members of one class rather than another. As a result, there will be borderline cases in which it is unclear whether a "significant overlap" has been achieved. At the same time, the verdict will be clear in many cases. FEAR, for instance, appears to clearly satisfy the

significant overlap condition. Whereas it is easy to find instances of "fear" that are not instances of FEAR, we would be hard pressed to find more than a handful, if any, instances of FEAR that are not instances of "fear". This is to say that the overwhelming majority of cases of FEAR are also cases of "fear". On the view I have proposed, this is enough to justify the orthographic identity between "fear" and FEAR.

I emphasize that it is quite possible that, despite including significantly more cases of "fear" than cases of non-"fear", FEAR will still comprise a very small number of cases of "fear". In other words, it is compatible with the rationale for orthographic identity I have offered that *most* cases of "fear" will *not* be cases of FEAR, as long as most cases of FEAR are cases of "fear". The fact that certain forms of "fear" are rarely instantiated is not a good reason for lacking a good scientific theory for them.

I leave it open for debate whether Panskepp's seven primary systems *all* satisfy either the subset condition or the significant overlap condition.

5.4 Heterogeneity is the problem

The real trouble with trying to build affective science around folk concepts lies in the "heterogeneity [of] the extensions of the terms scientists have borrowed from everyday language: *emotion, fear, anger, jealousy, mood,* and so on" (113). According to what I call the *Heterogeneity Hypothesis*, instances of folk emotion concepts are too different from one another to participate in the same body of scientific generalizations. If so, no general scientific theory of emotion, or theory of fear, or theory of anger will be forthcoming.

Russell has argued for the truth of this hypothesis over almost three decades (cf. Fehr and Russell, 1984). His view is that the majority of emotion theorists still deny – at least implicitly – the Heterogeneity Hypothesis. They believe instead that "[a]ll emotions are produced by (or are) a single type of entity (such as a set of affect programs in the limbic system). That entity produces (or is) an organized pattern of component responses" (81). This assumption extends to lower levels of the emotion taxonomy: all fears, all angers, all disgusts, all shames, all guilts are assumed to be produced by a single type of entity that generates the organized pattern of component responses characteristic of, respectively, fear, anger, disgust, etc.

This leads emotion theorists to keep formulating alternative accounts of the allegedly single type of entity that corresponds to emotion, or to fear, or to anger. The striking fact is that they keep failing at this task. For any theory that claims that all emotions are X, or that all fears are Y, or that all angers are Z, counterexamples can be found (Scarantino, in press).

The Heterogeneity Hypothesis provides what I consider to be the best explanation for this failure. If the hypothesis is true, the things that fall into folk emotion concepts constitute highly heterogeneous groupings, and there simply is no X such that all emotions are X, no Y such that all fears are Y, no Z such that all angers are Z, and so forth.

The Heterogeneity Hypothesis can be further supported by reflections on the origin of folk emotion concepts. Quite clearly, such concepts were not introduced in language to collect homogeneous domains of scientific investigation. Their function was rather to allow speakers to engage in an assorted variety of speech acts, which include describing and predicting states of affairs, making promises, issuing warnings and commitments, broadcasting intentions, offering praise and blame, expressing internal states, faking internal states, recommending or urging courses of action, and so on. It would be purely coincidental, and highly surprising, if the concepts that emerged from these multifarious social practices happened to collect a set of items that are interestingly alike from a scientific point of view.

I therefore conclude with Russell that "emotion [or fear or anger etc.] cannot be equated with any one process or component. Emotional episodes are not all of one kind, and they do not all stem from one mechanism dedicated just to emotion [or fear or anger etc.]" (82).

The question is: What follows from the heterogeneity of folk emotion concepts?

6. Why do emotion components co-occur?

Russell thinks that the heterogeneity of folk emotion concepts calls into question the standard understanding of the causal relation between emotions and their components. Consider the case of fear. "In the standard account", Russell writes, "fear (or the fear program) made Sally's heart pound, her palms sweat, her face broadcast danger; fear focused her thoughts on the bear; it made her freeze and then flee" (81).

The standard account here exemplified makes two central commitments. Firstly, it is assumed that there is a discrete causal emotion mechanism (or program) responsible for the co-occurrence of the various components (heart pounding, sweaty palms, freezing, etc.). Call this the *emotion mechanism* assumption. Secondly, it is assumed that the relevant mechanism or program is a "fear mechanism", which suggests that it is a mechanism common to all cases of fear. Call this is the *unique mechanism* assumption.

According to Russell, the emotion mechanism assumption and the unique mechanism assumption are both false. I will argue instead that, whereas the unique mechanism assumption is false, the emotion mechanism assumption is true. Now, the absence of a unique fear mechanism follows directly from the Heterogeneity Hypothesis, which both Russell and I subscribe to. One of the key respects in which members of folk emotion categories such as fear are heterogeneous is that there are no causal mechanisms common to them all.

However, the absence of a unique emotion mechanism is perfectly compatible with the presence of multiple emotion mechanisms corresponding to different instances of the folk concept. For instance, the fact that there is no discrete emotion mechanism that brings about fear components in all cases of "fear" is compatible with there being an emotion mechanism – call it the *fear* mechanism* for now – that made Sally's heart pound, her palms sweat, her face broadcast danger and her legs freeze and then run.

Russell's conviction to the contrary is grounded in an *argument from explanatory superfluity*. His view is that we can provide a satisfactory causal explanation of the co-occurrence of the components without positing any discrete emotion mechanism. My view is that the alternative explanations considered by Russell are not satisfactory. I will conclude that the best explanation for why the components of Sally's fear co-occurred is still the presence of a discrete – but not unique – causal emotion mechanism.

Russell's alternative "explanation begins with each of the components (her heart pounding, palms sweating, etc). Each component has its own causal chain, which research has gone a fair way in describing. Thus, we need an explanation of how the encounter with the bear resulted in changes in the autonomic nervous system, but that explanation would not include the assumption that fear did it. And so on, through each component that actually occurred during the emotional episode" (82).

This formulation appears to beg the question. It is assumed that each component has "its own causal chain", when the issue is whether or not there exists a "common causal chain" running from an internal emotion mechanism to each component. On the other hand, it is true that there is a cause – potentially the *same* cause – for each component. Furthermore, since Sally's fear consists of the *joint* occurrence of several components, what really needs to be explained is why the components co-occurred, not simply why each of them occurred independently of one another.[3]

3. Russell rightly emphasizes that in many emotion episodes the correlations among components are much weaker than in prototypical cases (e.g. Sally's case). A great many non-prototypical instances of folk emotion concepts occur by virtue of a few co-occurring components.

Russell is well aware of this explanatory need, and proposes three non-mutually exclusive non-emotional explanatory mechanisms: "(a) features in the environment have a correlational structure, which then creates correlations among components, (b) one component can influence another, and (c) two components are correlated when they are both influenced by the same central mechanism such as attention" (83). Let us consider these candidate mechanisms in turn.

According to the first, if two environmental features F_1 and F_2 are naturally correlated, and they independently cause, respectively, components C_1 and C_2, then C_1 and C_2 will correlate despite the absence of a discrete emotion mechanism causing both. Russell gives the following example: "suppose that novel events [F_1] are more likely than familiar events to block a goal…Suppose further that goal blockage [F_2] elicits an ANS pattern of cardiac acceleration [C_2] and that novel events elicit frowns [C_1]. If so, cardiac acceleration [C_2] and frowning [C_1] will be correlated even if no internal process links the two" (Russell, 2009: 1273).

According to the second mechanism, if a component C_1 causes a component C_2, then C_1 and C_2 will correlate despite the absence of a discrete emotion mechanism causing both. Russell gives the following example: suppose that making a threat expression [C_1] "alters breathing and muscle tension, which in turn alters ANS activity, perhaps cardiac acceleration [C_2]. The consequence would be that the threat face [C_1] is correlated with cardiac acceleration [C_2]" (Russell, 2009: 1273).

According to the third proposal, if some mechanism M other than emotion causes component C_1 and component C_2, then C_1 and C_2 will correlate despite the absence of a discrete emotion mechanism causing both. Russell gives the following example: "suppose that focused attention [M] produces both muscle tension in the face [C_1] and cardiac acceleration [C_2]. If so, muscle tension in the face [C_1] will be correlated with cardiac acceleration [C_2]" (Russell, 2009: 1274).

Let us now apply these three mechanisms to Sally's case. We could argue that the bear was a novel event [F_1] which caused Sally's fear expression [C_1], and that this novel event correlated with goal blockage [F_2] which in turn caused Sally's heart to pound [C_2]. According to Russell, this would be a candidate explanation for why fear expressions and heart pounding correlate. Furthermore, we could rely on the second mechanism, and suggest that making a fear expression [C_1] alters breathing and muscle tension, finally leading to cardiac acceleration [C_2]. This would qualify as a second candidate explanation for why fear expressions and heart pounding correlate. Finally, we could appeal to the third proposal, and argue that the attention mechanism M produced both the fear expression [C_1] and cardiac acceleration [C_2], offering a third candidate explanation for why fear expressions and heart pounding correlate.

The problem with these alleged explanations is that they fail to provide any understanding of why the correlations among components take the form they do. We are told which alternative mechanisms could in principle bring about the components of Sally's fear, but we are not told what underlies the causal relationships between mechanisms and components. What is left in the dark is precisely what needs to be explained. Why would the novelty constituted by a bear cause both heart rate acceleration – through goal blockage – and the facial expression of fear, rather than something else? Why would fear expressions cause heart rate acceleration, rather than something else? Why would attention shifts cause both heart rate acceleration and the facial expression of fear, rather than something else?

On the contrary, positing an internal emotion mechanism – what I have called so far the fear* mechanism – provides a clear functional explanation of *why* the correlations among components take the form they do. On this proposal, the reason why the novelty constituted by a bear leads to the co-instantiation of heart rate acceleration and a face expression and sweaty palms and freezing and fleeing behaviors – possibly by means of causal relations among the components themselves and attention shifts – is one and the same: all such changes are functional to the avoidance of danger and brought about in a coordinated fashion by an internal mechanism for that very reason. [4]

On Russell's model, the co-occurrence of the components of Sally's fear is *coincidental* rather than *functional*, because his three proposed mechanisms operate independently of one another and are not goal-oriented. Russell appears to explicitly endorse a coincidental explanation when he compares the co-occurrence of components of a prototypical emotion episode to the occurrence of a royal flush in poker – a combination of ace–king–queen–jack–ten of the same suit. As Russell remarks, "beyond the rules of the game and statistics, there is no need for an extra mechanism (except in cheating) to explain the very occasional occurrence of a royal flush" (84).[5]

But it is far from coincidental that being exposed to a bear is followed by the components of prototypical fear. Whenever we expose Sally, and a great many other animals for that matter, to bears, a very similar set of components will co-occur, over and over again. This suggests the presence of an internal causal

4. This is not to say that there is always a danger at hand when fear is elicited, nor that fear is always conducive to the avoidance of danger. The claim is simply that danger-avoidance is what the fear system is supposed to bring about.

5. Russell presents the even more radical analogy between emotions and the élan vital. In this case, Russell seems to lean toward eliminativism, because there is no such thing as élan vital. Discrete emotions do exist, on the other hand, and they are causally important for the reasons I describe in the text.

mechanism that *couples* dangerous stimuli with componential changes – the fear* mechanism. This is to say that the "very occasional occurrence" of prototypical fear – the royal flush mentioned in Russell's example – is not due to the laws of statistics, but rather to the fact that the fear* mechanism rarely encounters stimuli dangerous enough to demand a prototypical fear response.

So I conclude that Russell is right in rejecting the *unique mechanism* assumption, but wrong in rejecting the *emotion mechanism* assumption. Positing an internal emotion mechanism, I conclude, continues to be the best explanation for the systematic relation between classes of stimuli (e.g. dangerous stimuli) and classes of componential changes (e.g. heart pounding, palms sweating, fear expressions, freezing and fleeing).

This being said, I follow Russell in thinking that the standard account of the relation between discrete emotions and emotional components needs to be rejected because it posits the existence of a *unique* causal mechanism. Furthermore, I have only characterized the fear* mechanism functionally, without saying much of anything about how it is supposed to work. The next section aims to fill these gaps, and propose a reconciliation between Panskepp's and Russell's positions on emotional causation.

7. A possible reconciliation?

My central proposal is that the fear* mechanism at work in Sally's case is nothing other than Panksepp's primary FEAR system, in combination with learning and cognition. This hypothesis adds to the functional level of explanation an implementational level (Marr, 1982). Rather than simply considering the goal of the mechanism – e.g. avoiding danger – we are now considering a specific proposal on how such goal is physically realized.

On this proposal, Sally's response to the bear is driven by a genetically based subcortical brain network – the FEAR network – that coordinates sets of behavioral and autonomic outputs in ways that have proven adaptive in the evolutionary history of the species because they have helped animals "reduce pain and the possibility of destruction" (Panksepp, 36).

Now, Russell seems to accept that the FEAR network is involved in Sally's fear. According to Russell, "Panksepp's FEAR circuit…results in momentary freezing, inhibition of ongoing behavior, and changes in Sally's autonomic nervous system and endocrine system" (112). This admission is quite surprising, because these changes, unlike those accounted for by the three mechanisms I discussed in Section 6, *do* require positing an extra emotional mechanism – the FEAR mechanism.

But Russell adds that there are *other aspects* of Sally's fear episode that the FEAR mechanism cannot account for. This suggests that, rather than being *explanatorily superfluous* as posited in Section 6, the FEAR system may be *explanatorily insufficient*. For instance, Sally must perceive the bear, she must recognize it as a bear, she must produce changes in Core Affect in the direction of high arousal and high negativity, she must enhance perceptual-cognitive processing of the bear, she must search her memory for previous bear attacks, she must search the environment for escape routes, she must formulate a plan about what to do and so on.

I think that a reconciliation between Russell and Panksepp on whether discrete emotions are genuine causal mechanisms can be reached at this juncture. The reconciliation hinges on whether Russell can accept that the FEAR system has a global causal reach on the resources of the organism, and on whether Panksepp can accept that the FEAR system, understood along the lines I suggested in Section 5.3, has a great deal of variability in its operation.

Russell argues that Sally's fear episode includes changes that need to be explained in terms of perception, cognition and memory rather than in terms of the FEAR mechanism. But the FEAR mechanism as characterized by Panskepp has a *superordinate* control function with respect to *all* organismic sub-systems available to the organism, including perception, memory and cognition.

The job of the FEAR system is to coordinate an open range of resources by gating and modulating incoming sensory inputs relevant to behavioral and autonomic outputs and by engaging in reciprocal interaction with the brain networks involved in learning and higher thought, thereby bringing about secondary and tertiary affects. So while it is true that Sally's fear episode involves a great many changes that go beyond freezing and inhibition of ongoing behavior, I think such changes can be included within the explanatory purview of the FEAR system in interaction with other organismic resources.

On the other hand, Panksepp should accept that the FEAR network responsible for coordinating the components of Sally's fear is not an "important *part*[] of an intrinsically ambiguous conceptual *whole*[]" (3) named by the English term "fear". As emphasized by Russell, there are many cases of fear that have nothing to do with the FEAR network. I have consequently proposed that we interpret FEAR as designating a theoretically motivated coordination system – rather than a *part* of the vernacular notion of fear as suggested by Panksepp.

Furthermore, Panskepp's FEAR system should explicitly be understood as a homeostatic property cluster kind (see Section 5.3) with a great deal of variability in the combination of components that can instantiate it. I take the presence of variability to be the primary motivator behind Russell's rejection of the view that

all emotions, or all fears, or all angers, are produced by (or are) a single type of entity. We should co-opt this insight into theories of discrete emotions, and make room for mechanism-dependent variability.

Mechanism-dependent variability consists of the fact that the same causal mechanism can bring about different components in different occasions. This is to be contrasted with across-mechanisms variability, which consists of the fact that there are multiple causal mechanisms, each with its own range of mechanism-dependent variability, underlying episodes of the same folk emotion.

Since Panksepp's FEAR system has the avoidance of danger as its goal and dangers differ across occasions, we should expect that the FEAR system will bring about different combinations of components in different circumstances. For instance, if the bear is faraway, the FEAR mechanism should bring about freezing and some combination C of other components. But if the bear is suddenly looming, the FEAR mechanism should bring about fleeing and a different combination C' of components.

On this view, the variability emphasized by Russell is not only compatible with the presence of a discrete causal mechanism underlying the co-occurrence of emotion components, but to be expected insofar as such variability is functional to what the mechanism is trying to do – avoiding the specific danger at hand in the specific circumstances at hand.

This being said, I agree with Russell that the FEAR network as such cannot be credited with producing Sally's meta-experience of fear. This meta-experience requires focused attention and concept use, and it is more likely to occur when she is safely back home than when she is facing the bear. Yet, the absence of the meta-experience of fear is not tantamount to the absence of any subjective experience associated with a FEAR episode.

The subjective experience associated with a FEAR episode consists of the way it is like to undergo the activation of the FEAR network in the specific circumstances in which it is activated. Since such activation can involve different combinations of components because of mechanism-dependent variability, I reject Panksepp's view that the activation of any primary system is associated with a distinctive and intrinsic subjective experience. On the view I am pushing, the activation of the FEAR system will generate different, context-dependent subjective experiences, depending of which specific behavioral and autonomic outputs are being brought about.

8. Conclusion

I have argued that discrete emotions are genuine causal mechanisms with respect to the manifestations commonly associated with them. At the same time, I have agreed with Russell that discrete causal mechanisms cannot be found at the level of folk emotion concepts such as emotion, fear, or anger.

The key corollary is that the science of discrete emotions should stop characterizing its objects of investigation in folk psychological terms. Panksepp's strategy of using capitalized versions of folk psychological terms strikes me as a good way to go, provided either the subset condition or the significant overlap condition are satisfied.

I have suggested that a possible way to reconcile Russell's rejection of standard accounts of discrete emotions with the existence of causal mechanisms is to modify Panksepp's model to account for the kinds of heterogeneity that Russell has so aptly characterized. Whether this modification can be successfully achieved within the confines of Panksepp's original theory remains to be seen.

References

Boyd, Richard. (1999). Kinds, complexity and multiple realization. *Philosophical Studies, 95*, 67–98.

Fehr, Beverley, & Russell, James A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113*, 464–86.

Levy, Robert I. (1973). *Tahitians.* Chicago: University of Chicago Press.

Marr, David. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. New York: Freeman.

Pfaff, Donald. (2006). *Brain arousal and information theory: Neural and genetic mechanisms.* Cambridge, MA: Harvard University Press.

Russell, James A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.

Russell, James A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion*, *23*, 1259–1283.

Scarantino, Andrea. (In press). How to define emotions scientifically. *Emotion Review*.

Smith, Kyle S., Mahler, Stephen V., Peciña, Susana, & Berridge, Kent C. (2010). Hedonic hotspots: generating sensory pleasure in the brain. In Morten Kringelbach & Kent C. Berridge (Eds), *Pleasures of the brain* (pp. 27–49). Oxford: Oxford University Press.