Emergence and Empathy *[final version, now published in Institutions, Emotion, and Group Agents: Contributions to Social Ontology, ed. H-B. Schmid A. Ziv Konzelman. Dordrecht: Springer, 141-158]*

Ronald de Sousa

University of Toronto

sousa@chass.utoronto.ca

'I weep for you', the Walrus said:
'I deeply sympathize.'
With sobs and tears he sorted out
Those of the largest size,
Holding his pocket-handkerchief
Before his streaming eyes.

Lewis Carroll

The topic of empathy has recently received a good deal of attention, both for the questions it raises about its mechanisms and for the role it might play in motivating moral behaviour. The present paper addresses both of these questions in the light of considerations about how shared experiences emerge from emotional interactions between individuals. It comprises three parts. I begin with the more general issue of the reducibility of collective emotions to individual emotions or other, sub-emotional, individual states. I do this by briefly situating the discussion in the context of two notions that have been somewhat contentious in philosophy during the last few decades: externalism, and emergence. In the second part, I narrow my focus to empathy, and discuss some speculations about its mechanism. In a third and concluding section, I draw on both

the preceding discussion as well as some further considerations to buttress sceptical doubts about the importance of empathy for morality.

1. Emergence and Externalism.

If there are collective emotions in some interesting sense, they must consist of more than the mere conjunction of a number of individual emotions. But what could be meant by this "something more"? In a collective or compound phenomenon, what counts as being "more than the sum" or "merely the sum of its components" is notoriously difficult to define. These phrases stir the passions of defenders and opponents of "reductionism"—another contentious term—and touch on live questions about "externalism" and "emergence" in the philosophy of mind. Externalists place the locus of certain mental states outside the individual brains where common sense tends to place them. And if externalism accounts for some features of individual minds and bodies, perhaps it can do the same for individual consciousness. Perhaps the very nature of some emotions might best be understood in terms of facts outside the mind of the person said to be experiencing them. Even an individual emotion, then, might be what it is as a consequence of certain collective facts.

For an example of a claim about the emergence of a social phenomenon, consider this passage from a recent interview in *Wired magazine* of Nicholas Christakis, who has become something of a celebrity for his study of "social emotional contagion". Christakis notes that the average Facebook user has about 105 friends, and claims that those 105 "friends" do not influence your opinions and taste, whereas your (five or six) actual friends do. He speaks of such social effects as "emergent". In his explanation of that term, however, it is not entirely clear whether we are being given an analogy, another example of the very same phenomenon, or a distinct species of a generic concept of emergence:

> The example I give . . . is graphite and diamond. . . . The properties of graphite are completely different than the properties of diamond, and those properties do not reside in the carbon. They arise as a result of the patterns of interconnections between the carbon atoms. Therefore a group of carbon atoms can have different properties that have nothing to do with the carbon per se and have everything to do with the ties between the carbon [sic]. And that's what we're seeing about social networks. The same people assembled in different ways can give rise to different properties." (Zetter 2010).

Talk of "emergence" is a standard strategy for blocking the perceived threat of reductionism. But what is that threat, and why does it bother anyone? If having a certain atomic structure, say, H_2O, is necessary and sufficient for a phenomenal property such as liquidity to arise, does it mean that there is *nothing but* that structure in the world, and that the property in question is merely epiphenomenal? To worry about that question is to feel what I've called the threat of reductionism.

The objection to reductionism is temperamental rather than rational. Broadly interpreted, reductionism is no more and no less than the fundamental project of science. It rests on the methodological presupposition that explanations of complex phenomena are to be sought in terms of the nature, arrangement and interactions of their parts. But while it is often assumed that emergence and reduction are mutually exclusive, a clarification of those terms can show that assumption to be false: rightly understood, reduction is compatible with emergence. To show this, it is helpful to distinguish several grades of emergence, depending on the conditions, if any, under which facts about the base level *suffice to predict* those on the emergent level.

At level zero, the properties of a whole can be seen to follow immediately from those of its components. If I arrange four coins in a square, the pattern's property of squareness is immediately seen to follow from that arrangement of its components. Putting it this way merges

two ideas: the psychological obviousness of the pattern, which is a contingent fact about our ability to detect a given "Gestalt"; and the logical fact that the spatial relations among the coins *constitute* necessary and sufficient conditions to form a square. The psychological fact is commonly taken to attest to the logical fact, but neither entails the other. Some a priori or logical truths are not obvious. I might have to work hard to deduce the behaviour of the whole from the behaviour of the parts, even if it can be done using only the laws of logic and mathematics. In this case, although it would not normally count as a case of emergence, I might have trouble figuring out the deducible consequence. Theorems are said to be "contained" in their premises, but they are generally opaque to mere common-sense even when those premises are known. I therefore suggest that we think of this as a case of "epistemic emergence", where the inference is hampered by logical complexity. Call this, then, *Epistemic Emergence of level 1*. Obviously there is nothing troublesome about it.

Epistemic Emergence of level 2 is the most common case. Here mere logic and mathematics no longer suffice. Instead there are law-like connections between the phenomena at one level and those at the other. Typically, for example, there are bridge laws linking individual and global (or micro and macro) phenomena. These must be discovered empirically. Once discovered, they can serve to yield information about the behaviour of the whole on the basis of information about the parts. Thus I am told by those who know such things that many (if not all) the properties of a never-before-seen protein can be deduced on the basis of three lower-level facts: what amino-acids compose it, the "secondary" structure determined by the sequence of those components, and the "tertiary" or folding properties of the whole structure. That can be done only on the basis of laws or regularities that must be discovered empirically. If that were not so, it would imply, if it could be generalised, that all science could ultimately be done *a priori*. Assuming an irreducible need for empirical imput in such scientific matters, then, this represents a genuine level of emergence in relation to purely logical deducibility.

Epistemic Emergence of level 3 relates to level 2 in a way analogous to the way level 1 relates to level zero: in both cases, epistemic justification does not ensure psychological accessibility. Level 3 typically applies to deterministic chaotic systems. Prediction is a logical possibility; but in a chaotic system small differences in the values of some parameters can generate, after a sufficiently long time, arbitrarily large differences in outcome. In practice, these outcomes are therefore effectively unpredictable. Their surprising nature justifies calling them emergent, but again there is nothing there that seems mysterious or metaphysically incompatible with the spirit of scientific reductionism.

It is at the next step that we get the kind of emergence about which disputes arise and ideology makes claims for ontology. We get *Epistemic Emergence of level 4* when there is no logical possibility of prediction because the correlation between facts at the lower level and those at the emergent level can be established *only* on the basis of ad hoc empirical correlations. No independently established general laws and principles exist from which the specific relation between micro and macro levels can be inferred. This is what bothers people about the emergent character of consciousness in relation to its neural underpinnings: it's that we *can't imagine* how the latter can give rise to the former. Perhaps, at a more advanced stage of science, we will come to understand why consciousness *has* to exist when certain sets of neural connections are made in a certain sort of wetware. But is that not, after all, just like the relation between the atomic structure of carbon and the divergent properties of graphite and diamonds alluded to by Christakis? On that criterion, the fact that just this divergence of properties arises from just that difference in arrangements is equally mysterious. And yet, for reasons that remain unclear, most people worried about the supposed irreducibility of consciousness would remain unperturbed by Christakis's example.

Armed with this taxonomy of grades of emergence, bothersome or not, we can reformulate the question of the reducibility of a shared or collective emotion to the component events that constitute it. There are two ways to think of the grounding level. One alternative is to think of a

collective emotion as arising out of the emotions of the individual members of the collectivity. A more intriguing possibility is that the collective emotion arises not from individual emotions, but out of sub-emotional and sub-personal characteristics of the individuals involved. The analogy then would be with the way that, on some "component theories" of emotion, an individual emotion is ascribed on the basis of a number of components—behavioral, experiential, physiological, situational, and cognitive—which taken separately do not suffice to warrant the name of emotion. One might be led to that hypothesis by the notoriously bewildering characteristics of crowd behaviour, which seem to amount to more than the sum of the individual emotions of group members. Under certain conditions, it seems that a "psychological law of the mental unity of crowds" comes into play. Gustave Le Bon anticipated by a century the situationist point of view of John Doris (2002) when he noted: "It is only in novels that individuals are found to traverse their whole life with an unvarying character. It is only the uniformity of the environment that creates the apparent uniformity of characters.... All mental constitutions contain possibilities of character which may be manifested in consequence of a sudden change of environment." (Le Bon 1896, 7) That suggests that the individual emotions involved at the time a collective emotion is being manifested are not merely the constituents of that collective emotion, but are partly caused by the fact that each individual is a member of that particular group. That leaves indeterminate the mechanism responsible for producing the collective emotion as well as the individual emotion that are implied by it, allowing for the possibility that the type of basic interactions responsible are not in themselves (yet) emotions, but some sort of attunement of lower-level states. This might be comparable to the mechanism, whatever it may be, that commonly results in the synchronization of menstrual periods among members of a single household. In this last case, there may well be a kind of emotional harmonization that also takes place, but that emotional attunement is the result of a prior physiological attunement, rather than an emergent effect of the existence of individual emotions.

Recall that level 4 emergence is defined by the impossibility of deducing, from the examination of a single particle, however thorough, what properties will be generated when that particle is associated with other particles in a given configuration that has never yet been tested. This is precisely the point made by Le Bon in connection with the minds of individuals in a crowd. No amount of investigation of a particular individual's dispositions to behave in this or that way in isolation, including the dispositions such an individual might avow when asked about counterfactual situations, can yield a reliable prediction of what that same particle or individual will contribute to a group phenomenon.

This presents an intriguing analogy with the "problem of externalism" figuring elsewhere, in several versions, in the philosophy of mind and language. Hilary Putnam's claim that "meanings ain't in the head" was based on the fact that the reference of a term was not adequately fixed by any individual speaker's knowledge of its sense. I can accurately refer to an elm without having any idea what one looks like or how it differs from an oak. My reference is fixed not by my mental state setting up adequate conditions of recognition, but by the fact that elms and oaks are reliably identified by others who speak my language and on whose expertise my reference tacitly depends (Putnam 1973). This is a straightforward externalist thesis about meaning.

Sue Campbell (1998) has put forward a more contentious thesis applicable specifically to the identity of our emotions. She focuses on the predicament of Roxane, in *Cyrano de Bergerac*, who thinks she loves Christian, not just because he is handsome and brave, but because, as she falsely believes, he is the author of the fine poetic words actually spoken or composed by Cyrano. After Christian is killed, Cyrano's sense of honour stops him from revealing himself when Roxane insists she would love Christian even if he were ugly. (But then how does she know...? That love is rare indeed that does not "alter when it alteration finds". Our insight into counterfactuals is shaky at best, and particularly so where the counterfactual concerns emotions.) In the play's final scene, Cyrano asks to see Christian's last letter. He gives himself away by "reading" it aloud when it has become too dark for him to see it. Whom then does Roxane love? Campbell insists that it is

too late now to say that Cyrano is Roxane's true love: not because he is also now conveniently dying, but because the question has at least in part been decided against Cyrano, by Roxane's past actions over a long period of time: kissing Christian, marrying Christian, speaking of her love for Christian, mourning him for years.

Campbell's analysis is "externalist" in two ways. First, there is a genuine indeterminacy about the object of Roxane's love: it may have been caused or *prompted* by Cyrano's fine words, but it was *directed*, together with her letters, her kisses, and her thoughts, at Christian. The second point is both bolder and more subtle. It is the claim that to individuate a changing emotion requires an act of "collaborative individuation". We can see this as an extended form of Putnam's externalism about meaning: just as we don't have privileged access into the meanings of our own words, so also we fail to be the sole authority over the nature and object of our emotions. This thought might encourage an anxious awareness of the perils entailed by the power of others to define us, as well as a resolve to master whatever additional power we might claim over our own emotions by controlling the way we express ourselves. "To change emotionally," Campbell writes, "we appear to need situations to work through, and some history of success." (102). The point can be conservatively glossed in terms of classical learning theory: a habit is extinguished not in the absence of the stimulus. Extinction requires the presence of the stimulus coupled with the absence of the response. It's the response, therefore, that holds the key to change, and while by definition (some definitions, anyway) emotional expression is *involuntary*, that term is sufficiently elastic to allow for some more or less indirect control of at least some of our responses[1].

---

[1]To draw the starkest contrast with Campbell's thesis, one can turn to the purest form of existentialism, illustrated by Sartre's play *Les Mains Sales.* The central character, Hugo, a Communist Party member, had been ordered to kill a deviationist Party leader. He did indeed kill him, but out of sexual jealousy. The party having now adopted the dead leader's line, Hugo could save his life by admitting that he killed out of jealousy, not political conviction. But in the climactic scene of the play he chooses to be "non-recuperable" by the Party, by retroactively construing his motive as political. In this existentialist view, the psychological and collective facts don't matter. The individual can just choose by fiat the nature of his past act. (Sartre 1948).

Insofar as the identification of an emotion is a social rather than exclusively an individual fact, we have a specifically emotional version of externalism. This identifies two respects in which the collective phenomenon is not "merely the sum" of its individual components: one is that it is emergent, at what I've characterized as level 4, in relation to the individual emotions; the second is that the individual emotions themselves are in part defined by the collective context in which they appear. The individual emotion undergone by participants in a collective emotion depends in part for its very identity on what is happening outside of each individual.

2. Modes of sharing:

How then, in practice, might a collective or shared emotion be built out of multiple individual states? I begin by briefly surveying some standard ways in which two or more people can share an emotion. In speaking of "standard ways", I mean that for the moment I will confine myself to cases that presuppose no special group phenomena such as those alluded to by Le Bon, that is, no emergence beyond Level 3; but we shall soon see that Level 4 may become implicated as well. The basic cases I have in mind exemplify only three familiar patterns of causation: common, mutual, or reciprocal. Without any claim to be exhaustive, I distinguish joint attention, one-way influence, mutual influence, and purely epistemic influence.

(i) *Joint attention.* The capacity for joint attention emerges in the first six months of life (Butterworth and Cochran 1980). In the simplest case, joint attention might be merely coincidental. When two people are looking at the same thing then, if they approach the common point of focus equipped with similar background assumptions, very likely their responses to it will be somewhat similar. Such similarity could be due to an entirely general mechanism involving common knowledge, attitudes, assumptions, and perceptual capacities. Or it could be due to some specific prior agreement which leads the subjects to interpret the situation in similar ways, when that would otherwise be difficult or unlikely. (A convention establishing the meaning of symbols might be required, for example.)

In simple cases of this kind, the common feeling that results need involve no interaction between subjects. Interpreted in this way, the case of joint attention can be conceived as a species of a general type, where a single cause acts on several subjects. The common cause might not necessarily be that on which attention is focused. On the contrary, it might affect two or more people's moods and feeling without generating any awareness of its nature. A smell, for example, or worse an odorless chemical affecting the nervous system, might go undetected as such by those whom it affects, and yet result in shared moods of depression, anxiety, or panic. Or it might force itself into the consciousness of the subjects, but without being the object of intentional focus, as when, for example, several people are subjected to some unpleasant condition. Think of lining up for a show in the sun on a hot muggy day.

In practice, few cases will fall within this simple pattern of mere common causation. Unless neither is aware of the other, each will be influenced by the other. There is evidence that attention itself is not a purely cognitive phenomenon but involves emotional engagement. In the psychological literature about "joint attention", the term is reserved for cases where two or more people don't just happen to focus on the same object, but the common focus itself derives from a prior emotionally tinged engagement with one another. Thus Peter Hobson regards the capacity of an infant to engage in joint attention as the culmination of a three phase process:

> 1. The infant engages with someone else. 2. The infant engages with someone else's engagement with the world—and is 'moved'. 3. The infant achieves a new level of awareness that she is engaging with someone else's engagement with the world (in part through the process of being engaged with the other's engagement with herself). (Hobson 2005, 188).

In this kind of case, then, the mutual engagement is prior to the common focus, and jointness is a result rather than a simple cause of mutual engagement. If a certain emotional attunement is a precondition rather than a result of shared attention, we might have to concede that there is really

no simple case of joint attention (unless we mean to speak merely of cases where two unconnected observers just happen to be looking in the same direction). Genuine cases of joint attention are really more akin to the complex type (iii) described below.

First, however, let us look at an intermediate case.

(ii) *One-way influence* could take several forms resulting in the two parties experiencing similar, different or even opposite emotions (as in escalating antipathy). The case of empathy, of which more in a moment, is often of this sort, for the object of empathic feeling might not even be aware of the existence of the empathizing individual. In the most interesting, cases, however, which are also those most likely to generate unexpected consequences giving rise to ascriptions of emergence, the causation is not one way.

When all goes well, a one-way influence can result in a harmony between two people, manifested in similar emotional attitudes and attested by similarity of patterns of brain activity. So much is indicated by an experiment in which fMRI observations were made of the patterns of brain activity in a storyteller and that of a listener (Stephens and Hasson 2010). The authors found a remarkable correlation between the extent to which the listener understood the speaker's story as the latter meant it and the extent of overlap in brain activity. Another experiment showed that meaningful physical gestures of the sort involved in playing charades also gave rise to overlapping regions of activity in the various participants, linked to mutual understanding (Schippers, Roebroeck, Renkena, et al. 2010). In all these cases the agreement between participants is not necessarily emotional: all that the brain evidence is that something is shared at a sub-personal level linked to cognition. But if cognition, when shared, involves similar brain activity, there is no reason to expect that the same would not also apply to emotions.

Conversely, the absence of common presuppositions can be a serious obstacle. In a particularly demoralising piece of research, Brendon Nyhan and Jason Reifler have shown that when people have misconceptions, confronting them with evidence of their mistake can be

counterproductive, serving only to entrench their erroneous conviction (Nyhan and Reifler 2010). As Aristotle remarked in another connection, “when you choke on water, what will you wash it down with?”

(iii) *Mutual influence.* The most complex case, at least when just two people are involved, involves causation that goes in both directions. If the resulting experiences are similar in both individuals, we might get the slow dance-like emulation involved when two people are communicating in a harmonious way, whether in therapy (Charny 1966) or in ordinary conversation (Kimura and Daibo 2006). In such cases the outcome of the emotional coupling might be shared emotions; but it could also be, on the contrary, a growing estrangement that might still count as a collective feeling insofar as it is generated by the emotionally affecting interaction between the two participants. Once again, however, it must be noted that the original component individual states may not be of precisely the same emotional species as the collective fact we describe as ‘mutual estrangement’.

The dynamics of mutuality are highly diverse. Some of the diversity results from the fact that the feedback given and received by the participants might be positive or negative, and that it can involve sympathetic entrainment or antithetical entrainment. An illustration of the former is provided by Tom Nagel's Sartrean analysis of mutual seduction (Nagel 1979). On this model, each lover's desire is enhanced by perceiving the other's desire, which is simultaneously enhanced by perception of his own. The structure of desire, then, is analogous to that of reflections in a pair of facing mirrors. But there is a disanalogy, which is that the images in the mirrors get fainter, whereas on Nagel's model the desire is intensified by each reflection. This is a case of positive feedback, and since all cases of positive feedback are inherently unstable, no general prediction can be made about where it will end.

Where the feedback is negative, we can expect a stable equilibrium. An example of such an oppositional model is provided by the emotional phenomenon that Alain de Botton has called “Marxism”, in homage not to Karl but to Groucho, who disdained to join any club so vulgar as to

admit him as a member. Here the natural dialectic tends to foster contempt for anyone so undiscriminating as to fall in love with such an unworthy object as me: the less you love me, then, the more I can love you; but if both are subject to the same dialectic, the two will find an equilibrium (de Botton 1993, 53–64). In actual love affairs, even when there isn't the premise required by the Marxist dialectic idea, similar dialectics take place that sometimes enable a relationship to find equilibrium.

If mutual love is enhanced by positive feedback, on the other hand, the lovers can end up swallowing one another in a kind of dance that can end only in death, as in the legendary stories of Tristan and Isolde, or of Nagisa Oshima's *Realm of the Senses.* Scaling back, under the influence of negative feedback, means restraint, and constraints, and brings back Romantic love into the Classical fold of proportion and moderation.

(iv) *Purely epistemic influence*. By way of comparison with more general ways in which the mental states of various individuals can be brought into harmony, we should also, if only by way of contrast, note cases where a collective belief is generated by rational considerations of evidence. These needn't involve emotions. This is ideally what happens in the scientific community when someone publishes a convincing paper that establishes a fresh piece of knowledge, perhaps requiring most people to change their minds—showing, for example, that stomach ulcers are caused by a bacterium when orthodox medical opinion assumed they were due to stress. It is noteworthy that such rational conversions are in fact rare, precisely because people's previous convictions tend to be emotionally invested.

What then might we expect to be the forms of sharing more likely to apply to emotional states? Here emotion theorists generally agree that we should make a three-fold distinction.

(a) *Contagion* appears to take place without the intervention of any subtle cognition, or perhaps any cognition at all, in small babies as well as other animals. Contagion is reflexive, not reflective: it involves no effort of imagination or thought.

(b) *Empathy* appears to be something like a primitive perceptual state, differing from contagion in that it does not necessarily result in similar states in observer and observed.

(c) *Sympathy*, in which there is a more detached form of resonance between observed and someone observed to be in a given emotional state. As Jesse Prinz has put it, "Sympathy is a third person emotional response, whereas empathy involves putting oneself in another person's shoes" (Prinz 2010 forthcoming). This point can be slightly confusing for, as Prinz points out, the British moralists "used 'sympathy' in a way that is similar to the way I want to use 'empathy'." Sympathy is more clearly a cognitive phenomenon. It seems to be based on understanding the situation, or building a model of someone else's mind. xxx

Before discussing how empathy might work, and how it might serve morality and politics, let me note in passing that some forms of emotional influence discovered by recent research seem downright bizarre. There is evidence that we are susceptible to being influenced by our friends but "we are also beholden to the moods of friends of friends, and of friends of friends of friends—people three degrees of separation away from us who we have never met, but whose disposition can pass through our social network like a virus." (Bond 2008). This applies not only to moods but to obesity. According to Nicholas Christakis, whom I quoted above, the range of states that are subject to such transmission is surprisingly broad. It reportedly applies to "happiness and depression, obesity, drinking and smoking habits, ill-health, the inclination to turn out and vote in elections, a taste for certain music or food, a preference for online privacy, even the tendency to attempt or think about suicide." (Ibid.)

Many thinkers, including the early modern sentimentalists such as Hutcheson and Hume—and Mencius long before them, have regarded our capacity for empathy as fundamental to our capacity for morality. The capacity to share feelings that results from the basic faculty of empathy is undoubtedly important. But empathy is not directed equally to all. This becomes apparent, if we look at the patterns of influence that distinguish closer friends from more distant acquaintances.

So much, at least, is claimed by Christakis in the previously quoted interview with *Wired* magazine:

> Our friends' friends' friends affect us – meaning that there's a kind of social domino effect or a social contagion. Things ripple through the network and we can come to be affected, not just by what the people around us are doing, but by what people further away, that we don't even know, are doing. The best example of this is a children's game of telephone. You're the fifth in line, and the person whispers something in your ear that is erroneous. But it doesn't just include the errors that that person introduced. It includes all the accumulated errors of everyone else. So that's how we come to be affected by people downstream. (Zetter 2010)

Actually that seems a little strange, because in the kind of case they are talking about the influence is predictable, whereas the telephone game admits of random branchings—limited, to be sure, since one can't misunderstand just anything as just anything else—but still it's never a case of there being a definite probability of X turning into Y on the basis of the influence of Z.

Still, the analogy is interesting for that very reason. Given the diversity of temperaments among participants, the phenomenon of uniform causal influence—happiness increases happiness, depression deepens depression, etc.—demands an explanation. Even if a single person's mood has *some* effect on that of another, there was no a priori reason to expect the influence in question to be the same for all those affected. People can resist as well as endorse what other folks are thinking. In the case of emotion, however, it seems plausible that the influence should simply be based on imitation. What are we imitating? Well, there's evidence that posture is naturally imitated, as is facial expression, and on a broadly Jamesian view of emotion we would expect some feedback-type influence of that behavioral process on the participants' emotional state. But while imitation is a powerful strategy, it turns out to do best if not everyone else is doing it too (Boyd and Richerson 2005). This makes good intuitive sense: savvy investors know that contrarian strategies often work best. But that leads us to expect that not everyone is pre-wired for

imitation. And this fact, in turn, should prepare us for the possibility that mechanisms akin to imitation and empathy are unlikely to be central to the human capacity for moral response. Yet several authors have recently made just such a claim for empathy. I turn, then, to look a bit more closely at the possible uses of empathy.

3. Sceptical Thoughts on Empathy

It is consoling to imagine that even if we can't trust in God, Nature does everything for the best. Empathy, the etymology of which suggests that it implies "feeling as if you were on the inside" of another's experience, looks like one of Nature's best inventions: a shortcut to the motivation of altruistic actions. Mencius noted, twenty-three hundred years ago, that when you see a baby about to fall into a well you don't need to think about it before leaping to save it. That involves a kind of reflex-like response often mentioned in this connection; but it isn't actually obvious that it requires empathy. At that moment, the baby in question might not be feeling anything at all. Whatever we call it, the feeling that drives the response was not yet, for Mencius, a virtue: rather it was the emotional "root" of what, with proper education, becomes the virtue of benevolence ('ren'). Recently Jeremy Rifkin (2009), an author of blockbuster books on issues of public significance, and Frans de Waal (2009), a leading primatologist, have taken up the theme in a big way. The converse idea, that evil is the result of a lack of empathy, underlies a new book by Baron-Cohen (2011).

Rifkin and de Waal argue that humans are blessed with a capacity to respond empathically to one another's emotions. Both suggest that empathy has evolved for the benefit of humanity, along with our smarts, our language, and our love of kin. If we would only *realize* this, in both senses of the word, we could have what Rifkin refers to as an "empathic civilization": a whole new golden age.

There are a couple of reasons for the current interest in empathy, as global social glue or panacea for cultural conflict. One is that when globalisation brings mutually antipathetic value systems into direct confrontation, empathy promises a shortcut to mutual understanding. Empathising with another's pain, we earnestly hope, will automatically motivate us to alleviate it. Another is the scientific discovery of "mirror neurons" (Gallese and Goldman 1998), which seem to provide a mechanism, triggering the speculation that we have a neurologically guaranteed access to others' emotions and especially their pain.

Both these ideas are questionable. Not every brain scientist is convinced of the existence of mirror neurons in humans; but if they do exist it isn't clear that they relate directly to empathy. Mirror neurons light up when the motor system is activated, and they owe their name to the fact that they also light up at the sight of someone else doing the same thing. One can imagine different ways in which we might interpret this observation. The most obvious is that mirror neurons have evolved to facilitate imitation by simple observation. But the fact that a bunch of neurons are observed to light up under these two different sets of circumstances doesn't suffice to establish that hypothesis. In addition, only humans imitate from birth, and in humans mirror neurons remained conjectural until very recently. They were first clearly observed only in monkeys. But as Alison Gopnik pointed out, monkeys "don't actually imitate what other monkeys do: so the ubiquitous and powerful imitation we see in human babies can't just be there because they have mirror neurons" (Gopnik 2009, 207). Her observation about the difference between humans and monkeys remains valid despite recent evidence that mirror neurons do indeed exist in human brains as well (Mukamel et al. 2010). One form of such imitation arises at an amazingly early age: newborns are liable to pull out their tongue in response to seeing someone pulling theirs. As Gopnik interprets this, "this means that for babies imitation is both a symptom of innate empathy and a tool to expand and elaborate that empathy." (Gopnik 2009, 205). In itself, it's not clear why we have to assume that this necessarily implicates emotions or feelings, except in the simple sense of "the feeling of pulling my tongue out"; but insofar as we go along with the

Jamesian view that physical behaviour itself, as well as visceral responses, are reflected in subjective emotional feeling, it seems likely that the capacity for imitation will be linked to emotional experience: “Psychologists have shown that people unconsciously copy the facial expressions, manner of speech, posture, body language and other behaviours of those around them, often with remarkable speed and accuracy. This then causes them, through a kind of neural feedback, to actually experience the emotions associated with the particular behaviour they are mimicking.” (Bond 2008).

So what can we expect of empathy as a source of positive shared emotions?

In a forthcoming paper, Jesse Prinz has given several reasons for thinking that the answer is: Not much. First, he points out that we cannot assume—as do Rifkin and most of the commentators that have gushed about his book on the web—that empathy necessarily includes concern. That obviously begs the question in the context of a debate about the role of empathy as a necessary, sufficient, or even essentially relevant factor in the motivation of moral responses. Prinz also points out that “empathy in its simplest form is just emotional contagion: catching the emotion that another person feels.” As I observed above, contagion resembles a reflex more than an emotion. This means that if contagion counts as empathy, then a sophisticated act of imagination is not required for empathy; yet it is undoubtedly required for morality.

Prinz notes that our moral judgment does not generally track our empathetic responses: “for example one might charge that it is bad to kill an innocent person even if his vital organs could be used to save five others... Arguably, we feel cumulatively more empathy for the five people in need than for the one healthy person”. Actually, that might not be true, if we take account of Paul Slovic’s finding that people are more likely to respond emotionally to the picture of a single needy person without further discursive information than they are to the same picture accompanied by a caption pointing out that many other children are suffering the same plight (Slovic 2007). Prinz's point, however, stands. Prinz also points out that other emotions can be involved in generating moral judgments. In itself that need not point away from empathy, because

it seems to imply that empathy is itself an emotion: in fact, however, empathy is a capacity to resonate with somebody else's emotion, not an emotion in itself. It is not so much *an* emotion as a *window* onto another's emotion: the resulting experience can be of almost any emotion.

Prince makes another remark worthy of a small detour. “There are crimes against nature: such as necrophilia, incest, or bestiality. In these cases, the dominant emotional response is discussed, when the action is performed by another, and shame if we perform or even consider performing such an action ourselves.” Needless to say that depends on who “we” are. Some of us regard the very concept of a “crime against nature” as an offense against rational thought. To anyone who shares this view, Prinz's argument will seem feeble. But it raises a couple of interesting points. I myself am unable to empathize directly with the desire to perform any of the three sorts of actions Prinz alludes to; the idea of the first and third, at any rate, provokes a modicum of disgust (though the second leaves me indifferent). But I am equally unable to sympathize with the view that any of these behaviours are in themselves *immoral*. And in this case I find myself empathizing with precisely the people who commit such acts. I do so not in the sense of sharing their desire, but in the sense of feeling their hurt for the condemnation that their inclinations are liable to call down on them. This suggests that some forms of empathy, far from being simply contagion, result from a sophisticated selection that imagination is able to make between different aspects and levels of appraisal. As we'll see in a moment, there is actually some evidence for this from brain science.

There is an asymmetry between moral approbation and disapproval. Prinz suggests that “the sentiment of disapprobation” towards a kind of action is what a negative moral judgement amounts to. The asymmetry arises from the fact that, while approbation can indeed be directed at acts that one finds particularly admirable, most actions that are not immoral elicit no sentiment of either kind. A further problem with this proposal is that whatever empathy might contribute to moral judgment, there are many kinds of disapprobation and many grounds for it. If I disapprove of the aesthetics expressed in your choice of hats, that does not show that I find you immoral on

that ground. It is notoriously difficult to say just what is specific about moral disapproval as opposed to other kinds of disapproval. Prinz seems to me to be on firmer ground when he mentions the important emotions of guilt and anger, which Alan Gibbard has singled out as crucial to moral judgements (Gibbard 1990). The kind of disapprobation that involves anger or guilt is unlikely to pertain to matters aesthetic. So I endorse his conclusion that a sentimentalist theory can be based on such emotions such as anger and guilt, while empathy is of only minor importance.

As for the moral incompetence of psychopaths, it is not, contrary to the central thesis defended in (Baron-Cohen 2011), due not to a lack of empathy; rather it seems related to a more general deficit in their capacity to experience genuine negative emotions in response to present pain in others and even to the prospect of future pain for themselves. (Blair and Blair 2005). In fact, as we shall see below, there is no evidence that the psychopath lacks empathy. On the contrary, it is plausible to suppose that some of our power to hurt is due to the accuracy of their perception of other people's moods, emotions, and vulnerable points.

The fact that empathy is not in itself an emotion also provides good reason for Prinz's rejection of empathy as motivation. Motivation is entailed by actual specific emotions, rather than by empathy as such. In some cases, however—and these are the cases that have led people to think that empathy is developmentally important in small children—the appropriate emotions will be elicited in a child capable of empathy, while they may remain opaque to a child who is defective in that regard.

I will conclude by adding a few considerations that might strengthen Prinz's reservations about empathy, drawing on the main threads of the foregoing discussion. To begin with, empathy appears to be above all an epistemic tool: an avenue into the minds of others. It isn't evil psychopaths who lack empathy, but harmless autistic individuals. If empathy very likely evolved to yield insight into others' minds, that affords no particular reason to think its function is to make us nicer to one another. As Mark Rowlands (among others) has suggested, our primate

intelligence shows signs of having been designed above all to manipulate and outwit the competition (Rowlands 2009). Knowledge of others' states of mind is highly important in the pursuit of those “Macchiavellian” aims. The mechanical capacity for empathy seems to be just one tool, together with calculation, mirror neurons and perhaps direct mood contagion, in the arsenal of Macchiavellian intelligence. Like those other tools, it is just as likely to serve selfish ends as altruistic ones. Knowing how others think and feel is imperative for hypersocial beings such as we are, but it is no guarantee that we'll *care* about the people we're thus equipped to know about.

That empathy evolved not to make us nicer, but to make us better able to deceive, control, and manipulate, doesn't mean it isn't a good thing. Lots of good things, in evolution, have arisen as “exaptations”, mere side-effects of adaptations that originally had quite different functions (Gould and Lewontin 1979). (The delicate bones in our inner ear that enable us to parse music and speech started out as jawbones with which our crocodilian ancestors crushed their prey). But exaptations, by definition, weren't primarily shaped to fill that novel role.

In the case of empathy, one indication that empathy didn't primarily evolve for the sake of mutual aid is that there is surprisingly little correlation between feeling another's pain and being inclined to help. If you happen to dislike the person suffering, you can respond with glee, not sympathy. Brain studies confirm the distinction made above on purely conceptual grounds, namely that empathy is not emotional contagion (though the two are sometimes confused). On the contrary, unlike contagion, it is strongly modulated by attitudes. Painful experience endured by another person is viewed with remarkable indifference if that person is thought to “deserve” it. Vignemont and Singer (2006) have provided evidence that the triggering of empathic responses is modulated by prejudices and opinions and depends on our appraisal of the situation. And we don't need brain scanners to tell us that people can be entirely placid, or even enjoyably entertained, when witnessing the torments of some person or animal that isn't judged to be part of their crowd.

In fact, although empathy and compassion are commonly said to promote greater inclusiveness in our attitude to others, the truth may be precisely the reverse: in order to feel empathy, we must first regard someone as "one of us". All too often, empathizing humans are like Lewis Carroll's walrus weeping for the oysters he is gobbling. Without a prior commitment to doing good, feeling another's pain is just as likely to move you to give them a wide berth as to close in to help. If you're in pain, it's unpleasant to feel what you feel. If I allow myself to feel it, rather than just turn away or turn it off, it could be either because I already care about you, or because it is worth it for me to know what you're feeling, just so that I can be forewarned about what you might do.

Neither is it obvious that empathy is an indispensable condition of behaving morally. Kant was wrong in insisting that emotion should have no part in moral motivation. But even if motivation does require emotion, there are many powerful emotions that can move us to respond to the needs of others without empathy. You may want to help someone you pity; but pity isn't the same as empathy. You may want to help someone in need out of a commitment to equality, or fairness, without feeling the slightest empathetic resonance. In defence of noble ideals of freedom, you might fight for the rights of someone for whom you feel nothing but repugnance. You may be moved to fight against an injustice not out of empathy for the victim but out of indignation against the perpetrator. Or, to turn things upside down altogether, your righteous indignation might even be bolstered by empathy with the perpetrator rather than the victim: guilt at your kinship with the bad guys, together with shame at your very lack of empathy for the oppressed, has been known to out a check book more effectively than the next person's tearful compassion.

Jeremy Rifkin reminds us that "two and a half billion people in more than 190 countries watched the worldwide satellite transmission of [Lady Diana's] funeral... broadcast in forty-four languages... the most watched event in all of history." But what good did that do for anybody? None whatever. It may have been empathetic, but it was a fine illustration of sentimentality at its worst: what Oscar Wilde defined as wanting to have a feeling without paying for it. Paul Slovic has shown while a picture of a single suffering child might prompt people to give $20, they are

likely to give only $18 if they have to read any text with the picture, especially information about the many other children who share this one's plight. Their empathy works well enough, but a capacity for sober arithmetic might do a lot more good.

Boosting the possibility of moral progress are the indubitable advantages of collaboration, division of labour, the "win-win" strategies of trade and other cooperative undertakings. This prompts Rifkin to suggests that perhaps "human beings are not inherently evil or intrinsically self-interested.... and that ... drives that we have considered to be primary—aggression, violence, selfish behavior, acquisitiveness—are in fact secondary drives that flow from repression or denial of our most basic instinct" for empathic cooperation.

But it makes no evolutionary sense to suppose *anything* deserves the title of "most basic instinct". We are a patchwork of "modules" set up by natural selection to solve countless types of problems of living faced by our ancestors at all stages of evolution. These sometimes act independently and not seldom antagonistically, leading to the experiences of inner conflict noted by philosophers and psychologists ever since Plato decreed the soul to consist in three potentially warring parts. So we are all those things, good and bad, and many others besides. When Rifkin proclaims, in the face of the many instances of social collaboration afforded by complex modern societies, that "cooperation bests competition", he is oblivious to the self-refuting character of that slogan. "*Bests*" is a meaningless term outside a framework of competition. And the logic of natural selection is ineluctable: cooperation will indeed win out, if and only if it succeeds, at some appropriate level of selection, in *besting* competing strategies.

It's good to care about other people's pain and to be motivated to promote their welfare. We can all agree to that, but it won't usher in a new age. And while the capacity for empathy is one of the mental dispositions that sometimes might move us to promote social good, it is neither necessary in any particular case, nor ever sufficient in general. My money's still on the values of the Enlightenment: a little less stupidity, a little more passionate reason.

Conclusion

I began with a methodological plea for regarding different grades of "emergence" as reflecting differences in predictability from one level of analysis to another. None, I suggested, should be regarded as especially mysterious. When applied to collective emotions, this perspective leads us to expect that such emotions can arise as the sum of individual ones, but that they might also be causally grounded in sub-emotional, sub-personal physiological events, intensified and transformed by mutual causation. Among the forms of causation involved, it is often assumed that empathy plays a pre-eminent role, and functions as a crucial mechanism underlying our capacity for moral responses. I questioned both whether there is just one single mechanism underlying empathy and whether empathy should in turn be relied on to provide us with necessary emotional tools of moral response. In the course of making this argument, I distinguished three modes of causation pertinent to the states of two or more persons: simple common causation, one-way influence, and more complex forms of mutual causation in which reverberations can become indefinitely complex. In the more interesting cases of collective or joint emotion, involving all three of these levels of causation, the resulting collective emotion can be emergent at level 4. This means that it will not be possible to predict the nature of such a collective emotion on the basis of the properties of its constituents. The force of the externalist thesis I endorsed, when applied to the identification of shared emotions, is that the components of the collective emotions may not themselves be emotions. This provides a further reason why empathy—consisting either in simple contagion or in more sophisticated capacities for emotional understanding—will have no special role in explaining the *sui generis* collective emotion that emerges from the concurrence of individual phenomena. When a collective phenomenon results from complex interactions, not of merely additive individual emotions, but of sub-personal physiological and psychological states, implementing unpredictable and unstable causal processes, there will be no plausibility to the claim that the shared emotion is either justifying of or justified by the individual emotions. The bearing of such emergent phenomena on moral consciousness or

the disposition to moral behaviour seems bound to remain equally unpredictable, and we have very little reason to think it must be invariably benign.[2]

---

[2] I wish to thank conference participants for discussion at the 2010 Basel Conference at which these ideas were first presented, and I am particularly grateful to an anonymous reviewer for extremely helpful criticisms of an earlier draft of this paper.

References

Baron-Cohen, S. (2011). *The science of evil: Empathy and the origins of cruelty*. New York: Basic Books.

Blair, J., D. Mitchell, and K. Blair. 2005. *The Psychopath: Emotion and the Brain.* Oxford: Blackwell.

Bond, M. (2008, December 30). How your friends' friends can affect your mood. *New Scientist, 2689.*

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures.* Oxford; New York: Oxford University Press.

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. In L. Weiskrantz (ed.), *Thought without language* (pp. 5-25). Oxford: Oxford University Press.

Campbell, S. (1998). *Interpreting the personal: Expression and the formation of feeling.* Ithaca: Cornell University.

Charny, E. J. (1966). Psychosomatic manifestations of rapport in psychotherapy. *Psychosomatic Medicine, 28,* 305-315.

de Botton, A. (1993). *On Love.* New York: Grove Press.

de Waal, F. (2009). *The age of empathy: Nature's lessons for a kinder society*. New York: Three Rivers Press.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior.* Cambridge; New York: Cambridge University Press.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501.

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment.* Cambridge, MA: Harvard University Press.

Gopnik, A. (2009). *The philosophical baby: What children's minds tell us about truth, love, and the meaning of life.* New York: Farrar, Straus and Giroux.

Gould, S. J., & Lewontin, R. L. (1979). The spandrels of San Marco and the Panglossion paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, B 205,* 581-598.

Hobson, G. Peter. (2005) "What puts the jointness into joint attention?" In *Joint attention: communication and other minds ; issues in philosophy and psychology,* ed. Naomi Elian. Oxford University Press, pp. 198-220,

Kimura, M., & Daibo, I. (2006). Interactional synchrony in conversations about emotional episodes: A measurement by ''the between-participants pseudosynchrony experimental paradigm'. *Journal of Nonverbal Behavior, 30*(115-126).

Le Bon, G. (1896). *The crowd: A study of the popular mind.* New York: Macmillan.

Mukamel, R., Ekstrom, A., Kaplan, J., Iacoboni, M. and Fried, I. (2010). "Single-Neuron Responses in Humans during Execution and Observation of Actions". *Current Biology* 20(8), 750-756.

Nagel, T. (1979). Sexual Perversion (T. Nagel). In *Mortal questions* (pp. 39-52). Cambridge: Cambridge University Press.

Nyhan, B., & Reifler, J. (2010). When corrections fail: the persistence of political misperceptions. *Political Behavior,* 30, 303-330.

Prinz, J. (2010). Is empathy necessary for morality? In P. Goldie & A. Coplan, ed., *Empathy: Philosophical and psychological perspectives.* Oxford: Oxford University Press.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy, 73*(19), 699-711.

Rifkin, J. (2009). *The empathic civilization: The race to global consciousness in a world in crisis.* New York: Jeremy P. Tarcher/Penguin.

Rowlands, M. (2009). *The philosopher and the wolf: Lessons from the wild on life, death and happiness.* London: Granta.

Sartre, J.-P. (1948). *Les mains sales: Pièce en sept tableaux.* Paris: Gallimard.

Schippers, M. B., Roebroeck, A., Renkena, R., Nanetti, L., & Keysers, C. (2010, online before print May 3). Mapping the information flow from one brain to another during gestural communication. *PNAS.*

Slovic, P. (2007, 7 April). When compassion fails. *New Scientist,* p. 18.

Stephens, G. J. S., Lauren J., & Hasson, U. (2010, June 18). Speaker-Listener neural coupling underlies successful communication. *Proceedings of the National Association for Science.*

Vignemont, F. d., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Science, 10*(10), 435-444.

Zetter, K. (2010). TED 2010: Nicholas Christakis: Does this social network make me look fat? http://www.wired.com/epicenter/2010/02/ted-2010-nicholas-christakis-does-this-social-network-make-me-look-fat/ . Accessed January 14, 2011.