The Philosophical Significance of Neurological Research for Empathy

Abstract

The article inquires into the philosophical significance for empathy of the research on the mirror neurons, the related shared manifold hypothesis, micro expressions and the investigations that have grown up around them. Three of the consequences will be explicitly addressed. Evidence that such a capacity as empathy exists at all will be provided, but in the ironic spirit of proofs of common sense. In turn, the inquiry into existence will lead to the rehabilitation of introspection as a method of investigation proper to empathy, albeit with certain conditions and qualifications. Finally, the scope and limits of the shared manifold hypothesis, which conceptually implements the functional and causative role of mirror neurons, will be engaged. The result will be that the shared manifold is less transparent the more that it is shared. A bigger magnet will not make a difference.

The Philosophical Significance of Neurological Research on Empathy

Science has provided us with technology to make the activity of the human brain visible—rather like Galileo's telescope made visible the shadows of mountains of the moon and the satellites circling Jupiter. The analogy is a tempting one. Just as Galileo's telescope made visible mountains and satellites that had previously been hidden due to their distance and scale, likewise the functional magnetic resonance imaging (*f*MRI) apparatus enables examination of the "mountains" of the cerebral cortex and its activity, previously hidden by the interposition of the human cranium and neurology's micro scale. It is also tempting to say what the *f*MRI does *not* make visible is pain or the empathic experience of pain. The reply is direct. Oh, yes, it does. The *f*MRI data captures an image (say) with the anterior cingulated cortex of the brain heavily illuminated by activity, indicating neurological discharges that are correlated in a rigorous way with viewing vicariously a painful stimulus. You do see the pain—light goes on. This chapter will look for a way out of this impasse.

Thus, the argument of this chapter is that neurological research in empathy has philosophical significance. Three of the consequences will be explicitly addressed. It addresses doubts that such a capacity as empathy exists at all and puts them to rest. But it does so in such a way as to lead the philosopher to conclude that empathy is more transparent and "cognitively penetrable" than the science that affirms its existence. It

rehabilitates introspection as a method of philosophic inquiry especially used for empathy, albeit with certain conditions and qualifications.[1] Finally, it engages the philosophic coherence of the shared manifold hypothesis. This is already an ambitious meander through the field; and one topic that will *not* be engaged is the evolutionary significance of the research. My decision is not to tell another "just so story" or to marshal the comparative empirical data from multiple disciplines that would enable comparative tests to falsify a candidate adaptive hypothesis (and the space required for such an effort).

Regardless of the value of neurological research to philosophy or folk (social) psychology, the practical benefits of social neuroscience are extensive. The promise is great: restoring motor, psychological, and social functioning to persons who have experienced neurological brain damage. Medical intervention in the form of new therapies coming on stream such as the chemical activation of already existing but latent stem cells in the brain, stem cell replacement, or partial neurological regeneration are real, if not immediate, opportunities. So none of the statements made here, whether ironic or even critical, should be taken as invalidating the enterprise of social neuroscience. The moral task of eliminating suffering, which is also and primarily a human undertaking, is a task on which all are completely aligned.

Providing a Bottom to the "Bottom Up" Approach to Empathy

Simply stated, the neurological research is about the discovery of mirror neurons and the related research program that has opened up around the use of functional magnetic resonance imaging (*f*MRI) in support of it. This research initially showed that specific neurons in the primate brain discharged not only when the subject performed an action but when the subject watched the experimenter making a similar action, such as grasping a cup. These results have been extended progressively from motor control, to sensations (pain), to emotions proper, the latter as imaginatively reenacted since a large *f*MRI machine. The initial research on mirror neurons was highly invasive and used monkeys.

[1] A detailed drill down on the relational between empathy as vicarious introspection and introspection pure-and-simple is performed in Chapter ___ on Empathy and Introspection. However, for our purposes here vicarious experience is just another fine-grained micro experience. In order to avoid redundancy I will not repeat analyses performed in one argument and relevant to the other.

However, since that time different neuroimaging methods such as magnetic transcranial stimulation (TMS), functional magnetic resonance imaging (*f*MRI), and electro encephalogram (EEG) cartography enabled many of the results to be extended to humans. Evidence supports the conclusion that humans have mirror neurons, too, which are activated both when we take basic actions as well as when we observe the same kind of action in others.

The formulation of this scientific fact is an engaging story in itself and is part of the philosophic significance for empathy, which was a set of phenomena previously restricted to aesthetics, marginal interpretations of morality, and certain kinds of psychotherapy. Now we have the facts, at risk of only slight over-simplification, that as an individual is using empathy, the brain region F5 lights up and the brain's anterior cingualate cortex is visibly suffused with blood! Observing the simple forms of behavior captured in the experimental protocols involves neural region activation similar to those engaged during actual behavior production.

> "About 10 years ago we discovered in the macaque monkey brain a class of premotor neurons that discharge not only when the monkey executes goal-related hand actions such as grasping objects, but also when observing other individuals (monkeys or humans) executing similar actions. We called them 'mirror neurons'…Neurons [of the human brain] with similar properties were later discovered in a sector of the posterior parietal cortex connected with the area F5…"[2]

Further results show that specific areas of the brain are activated when the subject whose brain is being monitored by the *f*MRI machine is viewing pictures of hands and feet being painfully impacted by a door or a thumb being sliced by a knife (along with a cucumber). These areas correspond closely to those parts of the brain activated when humans actually experience pain (Decety/Jackson 2004; Jackson et al. 2005). Strictly speaking, mirror

[2] V. Gallese. (2007). "The Shared Manifold Hypothesis: embodied simulation and its role in empathy and social cognition," in *Empathy and Mental Illness*, eds. T. Farrow and P. Woodruff, Cambridge, UK: Cambridge University Press, 2007: 452. Philip L. Jackson, Andrew N. Meltzoff, and Jean Decety. (2005). "How do we perceive the pain of others? A window into the neural processes involved in empathy." *Neuroimage* 24 (2005): 771-779. See also J. Decety & P.L. Jackson. (2004). "The functional architecture of human empathy" in *Behavioral and Cognitive Neuroscience Reviews*, Vol 3, No. 2, June 2004, 71-100.

neurons were discovered via single neuron recordings whereas the *f*MRI has been one of the main tools used for exploring the neural correlates of empathy. But these are converging issues. Mirror neurons implement the capacity to communicate feelings (sensations and affects) at the level of embodied affectivity. Thus, mirror neurons provide the missing bottom to the "bottom up" approach to empathic receptivity. We have the capacity. Now it is a matter of showing it off by exercising it.

Prior to the discovery of mirror neurons, plenty of evidence was available that human beings lived rich, interactive socials lives. Obviously. However, what was missing was a neurological implementation of empathy that included an account of how those lives could be accessed. Empathy was a higher, executive function. This provided an opening for skeptics to dismiss empathy as a form of telepathy or a form of intuition based on inferences—educated guesses--from subtle behavior clues. The identification of mirror neuron has solved the problem of the missing implementation mechanism and enabled the deployment of empathy as a full, rich competence that founds human interrelations at large.

What then is it the existence of which is at stake? At a phenomenal level, the world and the human beings in it live in a context of joint affectivity. Individuals jointly attend to states of affairs in the world, and this includes how they feel about one another. Phenomenally and in everyday speech, this just means that I feel the way the other individual feels or not – I have sensations, aches and pains, hopes or fears, itches and pinches that qualitatively resonate with the other's experiences. We spend a lot of time and effort managing our feelings about others as well as ourselves. I appreciate what the other is feeling, because I feel it too. In some instances this "appreciation" gets asserted and checked and cross-checked in various ways and rises to the level of knowledge. In other instances, it remains less rigorous, and that is fine for purposes of everyday life, art, and ethics. Obviously, this does not necessarily happen all at once. As systems become more complex, history becomes more important. Parents, early caregivers, are devoted to ministering to the new born's body as regards nourishment, warmth, dryness, comfort, and emotional equilibrium. This is the matrix out of which empathy emerges – not only

for the caregiver but especially also for the child. The formation of structures of the individual's personhood to regulate emotional experiences depends on the availability of the empathic caretakers to sustain and restore equilibrium. We might say that Mom is creating a clearing for another person to show up. That clearing is her empathy with the neonate.

The process is a boot strap operation and becomes self-sustaining. As one individual gets to know another, the world appears to the one as it appears to the other. This phenomenon has the other as its object at the representational level. It is a representation of his feeling, which, in turn, is his the way the world phenomenally appears to him. It is a further step to appropriate the object of which his experience – as I experience it phenomenally – is an experience. For example, the content of the other's fear may or may not be available to me. I see the other is afraid. I experience his fear as fear, but I am not afraid of the snake, which he sees or imagines but of which I am unaware. Rather I experience fear as his fear as an attenuated vicarious fear. He is in danger and knows it and is afraid. I am in touch with his fear as fear but not necessarily of the object of which he is afraid. It is a further step that when the representational dimension is introduced, it means that I can now be mistaken – the appearance versus reality distinction opens up at the level of intentionality, which may be fulfilled by reference to an actual object I the environment or not.

The discovery of mirror neurons and related neurology has resulted in an explosion of interest around empathy and related forms of social cognition in the science, medicine, and the popular press. Thus, it is useful to consider that empathy would survive as a philosophically central capacity even if such a neurological research program were suddenly to be invalidated, as improbable as such a result may be. Prior to the discovery of mirror neurons, the approach to empathy was to consider it as a form of receptivity comparable to other nonstandard forms of receptivity found in other biological life forms. Thus, the ants of the north Sahara Desert were demonstrably shown to be able to navigate by means of ultraviolet light not visible to the human eye. When these rays were blocked, the ants wandered aimlessly. When the blocking filer was removed and the ultraviolet

light again made available, the ants moved purposefully back to their nest. [3] Similar considerations applied to echolocation used by whales or sonar in bats. This was different than what it was like to be a bat; but had rich points of comparison. Human empathy is a form of receptivity based on phenomena similar to vicarious experience and emotional contagion, but not reducible to them. To get to full human empathy, understanding and interpretation of an individual in relation to another had to be engaged in interpreting what was captured in empathic receptivity. This approach from the bottom up was illuminating and thought provoking but was limited by being an argument from analogy from one form of (possibly nonhuman) species receptivity to another. It was a bottom up approach without a bottom, since the form of receptivity was itself without any grounding independent of philosophic reflections on receptivity. Instead there was a useful collection of reminders around emotional contagion, vicarious experience, and alternative forms of receptivity.

Of course, this analogy was different from *the* argument by analogy to the existence of other minds. If global skeptical doubts were addressed, then empathy would be a useful tool for resolving local doubts about what exactly was the experience of other individuals in this particular situation. However, if the very existence of empathy as a coherent, non mystical, non telepathic, phenomenon were in doubt, then it would hardly be timely to propose that empathy was an ontological bridge between individuals, constitutive of human relatedness from the ground up. What was missing was the implementation.

This implies that a specific physical implementation makes it possible, with further effort, to posit the capacity corresponding to it. The physical implication—in this case, neurological implementation—unpacks what it means to have the experience. If I have the experience, then obviously I must have the capacity to have the experience, which, in turn, opens up the possibility of a whole field of experiences. For example, no one thought to doubt the capacity of seeing, given the proper operation of the eye and the visual cortex to which it is connected by the optic nerve. Even without the

[3] R. Wehner. (1976). "Polarized-light navigation by insects." *Scientific American* 235, No. 1 (July 1976): 106-115.

implementation details, the average person could open up his eye and *see*. The example of persons who were blind provided further confirmation of the general competence, since the missing capacity was individual. With a capacity such as empathy, up until the identification of mirror neurons, the matter was a tad more problematic. Of course, there was a philosophic history of applications of sympathy in ethics and empathy (*Einfühlung*) in aesthetics[4]. The evidence provided by phenomenology and psychoanalysis were also powerfully persuasive to those who studied those disciplines and appreciated the value of the methods and results. However, there was always skepticism from the perspective of a hard-nosed logical empiricist, analytic philosopher, or scientists with strong behaviorist leanings. Look at who was doing the reporting – phenomenologists and psychoanalysts! It is true that ordinary language philosophy bore witness as John Austin spoke of a "special counter-part feeling" and John Wisdom of extended introspection.[5] But these comments retained the status of parenthetical remarks, were little developed, and might be mistaken for non-existent telepathic powers.

The absence of a neurological implementation of empathy provided an opening for skeptics to dismiss empathy as a form of telepathy or a mere of intuition based on inferences—educated guesses--from subtle behavior clues. The identification of mirror neuron has solved the problem of the missing implementation mechanism and enabled the deployment of empathy as a full, rich competence that arguably provides a foundation for human social interrelations at large.

The amount of technology that goes into the creation of this scientific fact—that an individual knows the answer to the question "what is he feeling?" because the individual feels it vicariously—requires a tad more background. Subject A looks at a picture of a painful event and the area of the brain activated in Subject A is the same as when other subjects undergo painful stimuli such as immersing a hand for a half a minute in icy water or receiving a pin prick. In another instance, a neuron (a functionally related set of

[4] T. Lipps, *Aesthetik*. Hamburg: Leopold Voss, 1903: 120. David Hume. (1739). *A Treatise of Human Nature*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press, 1973. More on these two later.
[5] J.L. Austin. (1946). "Other Minds" in *Classics in Analytic Philosophy*, ed. R.R. Ammerman. New York: McGraw-Hill, 1965: 353-78 and J. Wisdom. (1946). "Symposium: Other Minds" in *Other Minds*, Berkeley: University of California Press, 1968: 206-229 - on "extended introspection" see 122, 137.).

neurons will be abbreviated as "neuron") in subject A is discharged when picking up a cup and a neuron in Subject B is likewise discharged as B perceives A taking the action of picking up the cup. The description of the computing facilities in which the numbers are captured, stored, and crunched to get to the level of an intelligible description of an event in a portion of the human brain—a significant increase in blood supply indicated activity is significant. Here is a sample of the technology necessarily interposed between the experiential input and the scientific fact that is output. The exclamation "Ouch!" is an exceedingly concise and powerful summary of megabytes of numeric data and hard to imagine transformations between data formats.

> MRI data were acquired on a 3-T head-only Siemans Magnetom Allegra System equipped with a standard quadrature head coil. Changes in blood oxygenation level-dependent (BOLD) T2* weighted MR signal were measured using a gradient echo-planar imaging (EPI) sequence) repetition time TR = 2000ms, echo time TE = 30 ms, FoV = 192 mm, flip angle 80°, 64 x 64 matrix, 32 slices/slab, slice thickness 4.5 mm, no gap, voxel size = 3.0 x 3.0 x 4.5 mm). For each scan, a total of 183 EPI volume images were acquired along the AC-PC plane. Structural MR images were acquired with a MPRAGE sequence (TR = 2500, TE = 4.38, fov= 256mm, flip angle = 8°, 256 x 256 matrix, 160 slices/slab, slice thickness = 1 mm, no gap)" (Jackson et al. 2005).

The point is to give the reader a sense of the complexity of the technology required to establish a scientific fact—Subject A is empathizing—"perceiving" to use Jackson's exact word--with the pain of the subject depicted in picture A. This data captured off of the Siemans *f*MRI machine specifically referred to that generated by individuals looking at pictures of another individual accidentally closing a door on his finger or slicing his thumb while cutting a cucumber. This description goes on for another two paragraphs which extend the discussion to the software such as MATLAB used to statistically smooth and analyze the data that is generated. It is not otherwise referenced in the article, and, presumably, will be useful if another team of scientists would like to try to replicate the results.

A NonStandard Existence Proof for Empathy: It Lights up!

The following is the proof of the existence of empathy. It is accompanied by additional background on the experimental set up that generates the confirming data.

The data captured off of the Siemans *f*MRI machine specifically refers to the data generated by individuals looking at pictures of another human accidentally closing a door on his finger or slicing his thumb while cutting a cucumber. Subject A in the *f*MRI is empathizing with the pain of the subject depicted in picture A. We know this because we get a read out of the brain activity occasioned ("caused") in the subject by this painful stimulus in comparison with other control pictures of neutral activities *via* the concurrent data capture occurring in the *f*MRI machine. This numeric data is translated into an image of the brain that is color coded with appropriate brightness as indicating an intensity of activity. The brain region in the subject is activated and the brain's anterior cingualate cortex is visibly suffused with blood while viewing vicariously the painful stimulus and empathizing with how it must feel. In the resulting image of the area of the brain, it lights up. Empathy exists. Q.E.D.

Naturally, this is not to be taken as an existence proof in the form of a classic syllogism with premises and a conclusion. At this point, the philosophically astute reader will wonder if I am joking or perhaps proceeding ironically. If anything, this is an argument such as that provided by G.E. Moore for the proof of the external world. This "proof" was actually a performance. He got up in front of his learned, serious colleagues from the British Academy, promising them a definitive proof and said, "Here is one hand; and here is another."[6] In the case of the application of *f*MRI to empathy, the light goes on! Does this perhaps "backfire" and mean that skepticism about empathy is justified? Far from it, Moore's proof was no less powerful for being unconventional. The work of social neuroscience has demonstrably driven the capacity for empathy in the direction of its neurological infrastructure and provided a foundation at the physiological level where previously only interesting and engaging anecdotes were available about emotional contagion. The one advantage that Moore's proof enjoys in comparison with the philosophical application of the work of Decety—he would not necessarily endorse what

[6] G.E. Moore. (1939). "Proof of an external world" in *Classics in Analytic Philosophy*, ed. R.R. Ammerman. New York: McGraw Hill, 1969: 81.

I am doing with his work--is that Moore does not require the interposition of a complex piece of technology—the *f*MRI machine--and the related necessary computer calculations. Even though the amount of calculation is probably modest—for example not requiring thousands of hours of calculations and therefore being surveyable by human oversight—the insertion of the *f*MRI machine to make visible the parts of the brain that are contingently hidden results in a certain opacity. The introduction of an *f*MRI machine into the proof changes the meaning of "proof" and so the "Q.E.D," as being proper to the classic form of proof that is directly surveyable by human insight has got to be deleted. This is an issue similar to – but different than – that created by the proof that four colors suffice to fill in a planar map.[7] The proof requires showing that no counter-example can exist. In turn, this requires checking over 1400 reducible configurations, a task that is practical only by using hundreds of hours of computing time on a high speed computer. In turn, the introduction of a computer raises issues about the possibility of a program bug (software defect), In both cases, the path of the proof lies unavoidably through a special purpose apparatus – in once case a standard digital computer, in the other case, a *f*MRI machine, arguably a different kind of computer. What we have philosophically is further evidence of existence that is part of a coherent network of social and scientific activity that forms the world of knowledge.

This approach has precedent. This invites comparison with Galileo's putting his eye up against the viewing optics of his telescope--a simpler device but a no less revolutionary apparatus than a *f*MRI apparatus--and seeing shadows of the mountains on the moon and the satellites circling the planet Jupiter just as (Galileo then inferred) Jupiter and the other planets circled the sun. In some instances, the learned Aristotelians who dared to look through the dangerous device literally did not see shadows or moons. They saw arbitrary patterns not otherwise intelligible. In other instances, the judicial authorities of the time refused to put their eye up to the viewing lens on the belief that the devil was literally inside of the tube transforming the images; and these devils might, for all we know, literally get inside the head of the viewer. It was not worth the risk. So they sent the

[7] Kenneth Appel & Wolfgang Haken (1977). Solution of the four color map problem, *Scientific American*, vol.237 no.4: pp.108-121, October 1977.

above-cited Aristotelians, who reported back that nothing meaningful was to be observed. The result, of course, was that Galileo was placed under house arrest until he recanted some of the theories that the evidence provided by the telescope was supposed to furnish. Obviously no such fate awaits Decety (or Gallese), since we would accept an invitation to the lab, take a lecture or two on statistical software, and compare the advice of different experts in the application to imaging of subatomic behavior of atoms in strong magnetic fields. Therefore, for purposes of this study, I shall continue to assume that the experiments function as designed and produce the data exactly as reported.

Empathy Through the "Lens" of the *f*MRI Machine

If experience has taught us anything, it is not to underestimate the power of technological innovation. The *f*MRI machine of the future may resemble that of today as little as Galileo's telescope resembles the Hubble Space Telescope. The *f*MRI might be miniaturized down to a helmet or a device the size of a Bluetooth ear piece. If that proves impractical, the same result— visibility to cerebral processes unconstrained by the size of the machine—might be produced by scaling up the *f*MRI machine to be the entire building with the subjects having a conversation with one another seated comfortably in armchairs. Advances in user interface design will make our relationship with the *f*MRI apparatus as simple as that of putting our eye up to Galileo's telescope and looking. In this case, the *f*MRI apparatus will be able to be logically deleted from the proof and we will be back at the level of folk (social) psychology along with Moore's application of common sense. That is in effect what happens when NASA posts pictures on the web from Hubble for our viewing.

The point is that we will need to learn to read what is reported by the *f*MRI in various scenarios just as Galileo's contemporaries had to learn to see the patterns visible through the lens as moons and shadows or mountains while squinting with one eye closed. Of course, the difference is that in the one case we are viewing physical objects, albeit ones vastly distant in space (and time), and in the other case we are viewing—what?

Howsoever we describe the target, empathy, as represented through the "lens" of the *f*MRI, we are at risk of a category mistake—using the social (or mental) to describe the neurological or vice versa. This is precisely because we are trying to navigate between at least these two different domains of organizational dynamics—the domain of people interacting socially and the one of neurologically relevant aspects of their effects on one another. I experience and exclaim "Ouch!" as I see a picture of the painful stimulus--and the area in my brain lights up.

The exclamation "Ouch!" is an exceedingly concise and powerful summary of megabytes of numeric data and complex statistical transformations between data formats. In every scenario, the challenge of interpreting the vast quantity of numerical data provided by the *f*MRI machine in terms of an actual human experience—pain, fear, explicative, etc.—remains the same. We have to learn to substitute this data file and its representation as a lit up image of the anterior cingulated cortex for the experience that causes me spontaneously to exclaim "Ouch!" It recalls the challenges in translating a poem between two different languages—nothing is lost in translation—nothing except the poem itself. What does this mean and where can we turn for help? What about the role of introspection?

The way out of the impasse noted in the opening paragraph of this chapter—the philosophical point--is that the neurological data becomes scientifically relevant when they are correlated with the self-monitoring and attentionally guided introspective reports of the subjects in the experiment. The report of the subject in the *f*MRI machine in Decety's lab as to the intensity of the pain is a part of the experimental setup. The pain becomes "visible" as the scientist lines up the reports of the experimental subject and the data read out. The *f*MRI data is not a representation of some private event, logically accessible to only one person, but is rather contingently available directly to the subject of the experiment from a first-person perspective as well as from a third-person one.

The subject can shift from first-person reporting to viewing the resulting image of his or her own anterior cingulated cortex (e.g.) as can any third-person. This empirical data is

publicly accessible as indicated by the read out (or color-coded image). What gets lit up by *f*MRI is an event in the physical world, contingently hidden from visual inspection by the human skull and the lack of ability of the eyes to detect the relevant blood oxygenation perturbations in specific neural areas.

Also relevant is the intensity of the pain that is reported and correlated with what was occurring in the brain as captured by the *f*MRI (Jackson 2005: 774). Of course, here the "report" is based on the awareness of that to which the subject was attentively monitoring as experienced introspectively. Further note the control case: the absence of pain from the subject's experience is also essential to the result, since the relevant area of the brain does not light up. Finally, the experiment distinguished different brain areas in response to the request to imagine how the subject would feel if he were in the picture versus how the person in the picture felt. This again highlights the difference from inside the *f*MRI between the self in pain and the other in pain, or, stated in a more pronominal form, the first-person and the third-person points of view. Different areas light up.

The power of the experiment (and its limitation) comes from precisely the correlation between the two levels of discourse, the folk psychological and the neurological. In fact, within this research paradigm and the design of distinctions relevant to empathy, every piece of the functional architecture of empathy can be lined up with a neural correlate of empathic functionality. In "The Functional Architecture of Human Empathy" (Decety & Jackson 2004), the authors finds neural correlates of consciousness (NCCs) for every component of empathy—shared neural representations, self-awareness, mental flexibility, emotional regulation. These four macro components form Decety's definition of empathy. (More on it below.) Granted, in some cases the correlate is relatively fine-grained whereas in others it is still coarse-grained; the progress already made should provide confidence that matters can be further driven down in the direction of the underlying cerebral hardware. The power of Decety's account of empathy is that it integrates both the receptivity and the understanding. From a folk-psychological point of view, we distinguish first- and third-person perspectives, the self-other, and controlling one's own emotions. The shared manifold is *not* available at the level of folk psychology,

which is why (as we shall see) it becomes the fulcrum that moves the explanatory apparatus. A multi-component definition of empathy at the level of its functional architecture acts as a strong counter to any perspective that would assert empathy is a single module the absence of which alone is responsible for autism, psychopathology, etc.

In addition to the above cited research, the distinction between imagining what an individual—for example, I would feel if I were in the other's place and imagining what the other is feeling is shown to recruit different areas of the brain, providing an underpinning for the distinction between first-person and third-person experiences. Coarsely-grained NCCs are available for the executive ("high-level') brain function(s) in tamping down or enhancing emotional responses. This is significant for getting to a full, rich application of empathy. For instance, if my reaction to the distress of another individual at the death of his father evokes in me distress at the real or imagined loss of my own father, then I will be flooded by an egocentric sense of loss. My own ego gets in the way of my empathy. The individual successfully using empathy to understand another has to balance his or her own loss to inform an appreciation of the experience of the other without, however, allowing it to cause distress and distract him from the other.

Simulation Machines "All the Way Down"

The operation of mirror neurons implies that humans are simulation machines "all the way down." If not "all the way down," then at least at a low, cellular level. In other words, the social world is simulated by us at a level that Collingwood could hardly imagine with his early, incisive account of imaginative reenactment and rethinking the thoughts that the Roman emperor Hardian thought.[8]In short, some simulations are below the level of attentional availability for introspection; that is, below the threshold required to attain consciously awareness. In contrast, other simulations at the level of self-

[8] R.G. Collingwood. (1946). *The Idea of History*. Oxford: Clarendon Press, 1968: 282.

reflexive decision can still be intentionally driven by imaging "what if" scenarios in planning and anticipating future outcomes.

By setting a limit to introspection—one cannot introspect the operation of mirror neurons though one can see them discharge on the *f*MRI —neurology has made it possible to think both sides of the boundary between phenomenal experience as available to attentional self-monitoring and the neurological infrastructure. This opens up a logical possibility, albeit a limited one, for the rehabilitation of introspection. In social terms corresponding to folk (social) psychology, what is available for introspective attention is the rich and diverse field of vague and ambiguous phenomena described as "vicarious feelings," "inner imitation," various fine distinctions of affect and sensation, as well as consciously driven decisions or "what if" scenarios. (This list is not complete.)

A "simulation" uses a different media to implement a function between a source and a target environment or scenario. For example, Kelvin's simulation of the tides in the ocean used ropes and pulleys—a totally different media—to represent the behavior of the tides. Without wishing to split semantic hairs, a simulation used to imply the media were different; otherwise, the target environment would be a copy, perhaps at a different scale or intensity, not a simulation of it.[9]

However, nothing depends on the specific use of "simulation" being in a different media, and the term may plausibly be extended. With mirror neurons we are using the same media and at the same size scale too! The same neurons are multifunctional, discharging when experiencing pain and when viewing another experiencing pain. Well and good. From another angle, we could assert that the mirror neurons simulate the other's experience of pain as I have a vicarious experience of the other's pain. Here we cross levels of organization from the relatively well-defined domain of neurology to that of the dense, interactive "causal thicket" of several interrelated domains -- human interrelations, our phenomenal experience of the other, the emergent properties of representation, vicarious experience, intentionality, and the meaning, including linguistic meaning, with

[9] K. Craik. (1943). *The Nature of Explanation*. Cambridge: Cambridge University Press, 1967: 51.

which we human give an account of our experience.[10] If one organism uses the same set of neurons to experience pain and to experience vicarious pain in watching the other individual experience pain, then "same" must be defined carefully. Here "the same" means "isomorphic," not numerically identical. Staying at the level of neurons, "simulate" means "copy," "repeat," or "imitate," albeit with a different level of intensity.

"Inner Imitation" as the Downstream Effect of Mirror Neurons

It is a common place that individuals do not perceive the discharge of mirror neurons – or any neurons for that matter. This results in a search for the downstream phenomenal effects of the discharge. One example is asserted to be "inner imitation," the mirroring of the bodily position of two people having a conversation. The one person leans inward; then so does the other. A photograph of an athlete preparing to let go with a bowling ball, hurling forward a heavy object that results in his leaning forward also shows the entire crowd leaning forward.[11] The photograph shows the crowd leaning forward so it is a further question in what sense the imitation is "inner." It is "inner" – or more properly speaking coded as "inner" – since the individuals are unaware that they are unwittingly imitating the bodily position of the athlete. In this instance, "inner" is contingently unaware, but available for introspective awareness if one's attention is called to it. If the individual were suddenly asked how he felt – say by the beeping arrival of an instant message – then the individual would fine the muscular sensations of leaning forward available to his introspective awareness. The latter is "inner" in the sense that it is accessible to a first-person report. This opens up a chase after an entire spectrum of fine-grained sensations and affects that are contingently available to one individual but not necessarily to another.

Probably the most famous—one might say infamous—proponent of inner imitation as the basis for empathy was T. Lipps. With the discovery of mirror neurons, Lipps is enjoying

[10] "Causal thicket" is a term coined by W. Wimsatt. (1994)."The ontology of complex systems" in *Reengineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press, 2007: 200.
[11] This photo graces the cover of E. Hatfield, J.T. Cacioppo, and R. L. Rapson. (1994). *Emotional Contagion*. Cambridge, UK: Cambridge University Press, 1994: 34 where the caption says "The All-Ireland Roads Bowl Championship."

a certain degree of *ex post facto* validation. Was Lipps introspectively aware of something which most other researcher since him have overlooked? Lipps writes that "empathy is nothing other than the inner aspect of imitation."[12] Lipps gives the example of contemplating an acrobat. This is aesthetically relevant to Lipps, for whom the human form is a paradigm of the beautiful:

> In inner imitation there is no separation between the acrobat up above and me down below. On the contrary, I identify myself with him. I feel myself in him and in his place. . . .
>
> In such inner imitation I discharge—not the motions which the acrobat discharges—rather I discharge immediately, or inwardly, or in thought, the motions of the acrobat. I discharge the motions in so far as this discharge of motion is not an external but an inner act in the acrobat himself (Lipps 1903: 121f.).

In order to make the connection between the visual aspect of the acrobat's performance and the viewer's introceptively available, kinesthetic feeling of soaring, tumbling, etc. Lipps identifies "an original innate association between the visual image and the kinesthetic image" that he declares to be "primitive" and "not further explainable" (Lipps 1903: 116, 118). Not further analyzable because it extends beyond the threshold of awareness or, alternatively expressed, does not rise to the threshold of awareness, whether sensory or affective. If one watches a man balancing on a high wire above the ground, one gets a sense of leaning back and forth, left or right, in resonance with what the acrobat is doing. Mirror neurons have instantiated what Lipps presciently envisioned in 1903 in the face of significant skepticism, we might add. Did Lipps have the introspective insight to attentionally perceive his owned sense of balance; did he have access to introspective experiences that are not available to most of us (all of us); did he have powers of imagination most of us lack; or was he simply a good guesser, a good speculator? These choices are not exclusive

It might not be the actual discharging of mirror neurons but perhaps an intermediate level representation of a common coding mechanism, a trace of which has percolated up into attentional availability? This would be slender phenomenal traces of inner imitation that

[12] T. Lipps. (1903). *Aesthetik*. Hamburg: Leopold Voss, 1903.

might be identified and reidentified with training such that they can become the basis for a certain delicacy of empathy, where "delicacy" refers to the ability to make fine-grained discriminations of vicarious feelings and affects rather like the perception of musical tones improves with training. Perhaps Lipps had the equivalent of "perfect pitch" for what is called "inner imitation," but then proceeded to build his genuine, initial insight into a wrong-headed theory of projective empathy, given that he was committed to explaining how aesthetic perception generates the experience of natural beauty.

This goes a long way to clarifying what has been dimly recognized but not well articulated about the relationship between aesthetic appreciation and human empathy.[13] The cultivation of an appreciation of beauty in its various forms--whether music, fine arts, or enjoyment of natural beauty—enhances an individual's empathic sensitivity. The development of an individual's empathy fosters an appreciation of music, artistic creativity, humor, and wisdom in the old-fashioned sense of appreciation of life's lessons learned.[14] What both the empathic individual and the aesthetic one share is a capacity for fine distinctions of feeling. Another name for this in aesthetics as theory of beauty is "taste." In short, taste is the capacity for judging the beauty of something by means of the feelings aroused by the object. It is particularly important, though not directly relevant to this thesis, whether this arousal precedes or follows the intellectual or emotional content in the mental act of accessing or capturing the beautiful object.

[13] David Hume. (1739). *A Treatise of Human Nature*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press, 1973. The distinction between impressions of sensation and reflection contains the nucleus of an account of "emergent" qualities. A sensation of pleasure or pain is embedded in an impression of reflection (desire, aversion or pride and humility – direct or indirect passions), but the latter is not reducible to the former without remainder. Hume reminds us that ". . . The minds of men are mirrors to one another, not only because they reflect each others emotions, but also because those rays of passions, sentiments and opinions may be often reverberated . . ." (*Treatise*: Bk II, Part II, Sec. V). It is perhaps because Hume assimilates to "delicacy of taste" the capacity for friendship and the appreciation of character that we fail to find anything like a "delicacy of sympathy" called out explicitly by Hume (who, as a rule, uses "sympathy" for what today we more-or-less call "empathy").

[14] H. Kohut. (1971*). The Analysis of the Self*. New York: International Universities Press. 1971.

Hume as Witness to the Intersection between Taste in Beauty and Empathy

What both the empathic individual and the aesthetic one share in common is a capacity for fine distinctions of feeling (sensations and affects). The name for this in aesthetics (the theory of beauty) is "delicacy of taste."[15] "Taste" is the capacity for judging the beauty of something by means of the feelings aroused by the object. The name for this capacity for fine distinctions of feeling in folk (social) psychology is "empathy." Many individuals, including experts dealing with hard cases, deploy attentional self-monitoring to make these fine-grained distinctions.

Once again, a particular philosopher, David Hume, is of interest as a witness to the intersection of introspection and human empathy. The relevance of neurology will follow shortly. Hume assimilates all the advantages for human interrelations of his widely quoted theory of "sympathy"[16] to "delicacy of taste." A brief note on terminology is required since, for Hume, "sympathy" overlaps substantially with what we call "empathy"; and "delicacy" is an eighteenth century term for the ability to make "fine-grained distinctions." Access to empathy is primarily through aesthetic taste; and the advantages that today would be attributed to "empathy," according to Hume, include an increase in a person's capacity for love and friendship and enjoyment of the "characters of men." The abilities that make men more sociable in the sense of being able to make enduring friendships come under "delicacy of taste," while all the disadvantages of increased sensitivity (in the sense of being easily upset, irritable, choleric) come under "delicacy of passion" (Hume 1741: 25-28).This is not wrong; it is just a difference in terminology, given a distance of two and a half centuries. This is one reason why no separate analysis of "delicacy of sympathy," Hume's term for what we today call "empathy," is found in Hume.

---

[15] David Hume. (1741). "Of the delicacy of taste and passion" in *Of the Standard of Taste and Other Essays,* Indianapolis: Bobbs-Merrill: 1965: 25-28. See also D. Hume, "Of the standard of taste" (1757), in *Of the Standard of Taste and Other Essays,* Indianapolis: Bobbs-Merrill: 1965: 3-24.

[16] Hume (1739) reminds us that ". . . The minds of men are mirrors to one another, not only because they reflect each others emotions, but also because those rays of passions, sentiments and opinions may be often reverberated . . ." (*Treatise*: Bk II, Part II, Sec. V

Any special capacity to feel—whether in taste or empathy--is dependent on the ability of our sensory organs to perceive the fine details of a composition and to further process this input introspectively. Let's consider an example of taste. Our taste is not tested in distinguishing (e.g.) a jagged, irregular configuration from a smooth, curvaceous one as clumsy from graceful. In this case, virtual insensitivity will do.[17] Arguably taste is required to make such a distinction, but it is not a good test of taste, for example, enabling us to judge that your taste is better than mine. It is just too coarse-grained. What is needed to activate our taste is an example where there is such a multiplicity of details that the capacity of our organs to perceive fine and exact ingredients in the composition is challenged. A closely related question is whether another can notice in his introspective attentional monitoring all the fine details to which I introspectively attend and yet still be insensitive to the feeling aroused by them. After ruling out inexperience, prejudice, or organic disturbance, we would be justified in saying with regard to his delicacy of taste, the other simply hasn't got it (Hume 1757: 11).

> . . . If the same qualities, in a continued composition and in a smaller degree, affect not the organs with a sensible delight or uneasiness, we exclude the person from all pretensions to this delicacy.[18]

We both notice the same things – we have the same sensations – but he has some limitation, perhaps an impairment, in his capacity for introspection. The field of my sensations available for introspective attention is richer or deeper than the other.

Hume gives a wonderful example where "taste" is meant literally as referring to fine-grained sensations on the tongue, not aesthetically. A connoisseur ("expert") is judging a cask of amontillado sherry and remarks that it is very good but has a slight taste of iron and leather. This is regarded as curious if not absurd until the cask is drained and a rusty key on a leather thong is discovered. The connoisseur had perceived a fine-grained, micro

[17] Ted Cohen makes this point in arguing against a distinction between aesthetic and nonaesthetic concepts based on taste. His conclusion is a Kantian one – taste judges, not by means of concepts, but by feelings. "Aesthetic/Nonaesthetic and the Concept of Taste," *Theoria* 39 (1973): 113-52.
[18] Hume, Standard of Taste, p. 11.

trace of a quality that was missed by the others. The "delicacy of taste" of the expert provided more fine-grained details of sensation of which the others were unaware.

The application to empathy (and the *f*MRI) is direct. Is empathy really required to see that another is enraged when he denounces and insults everyone and throws a tantrum? As indicated above with taste, in the final analysis, empathy is indeed required to identify a gross instance of anger; but it may seem otherwise since empathy is not properly put to the test by such an instance. Another instance with finer details is required before empathy can be activated in any interesting way and put to the test. That is the point, as indicated above, of the example where an expression of rage masks self-contempt.

Hume made extensive use of introspection and it is a source of both strength and weakness in his philosophizing. Of course, a limitation, if not weakness, was Hume's implicit belief that introspection could see "all the way down." He had no idea of mirror neurons, or any neurons for that matter. However, in spite of this limitation, Hume makes advances in the deployment of introspection in a controlled, regulated way, that are equal to any phenomenologist and have not been surpassed even today in any area. That is especially so in the perception of art and taste. These are lessons easily transferable to empathy. For example, if I sense the self-contempt of a person throwing a temper tantrum where everyone else only notices his rage, then a distinction only possible through a greater "delicacy"—ability to make fine-grained distinctions--of my empathy is made evident. In Humean terms, while all of us get the same sensation—hear the swear words, terms of abuse, see the stamping of feet—only one of us (in this case) is aware in introspection of the person's sadness and self-contempt. I am receptive to something to which you are indifferent and unaware. My "delicacy of empathy" is more receptive than yours; and whereas you perceive the rage, upon further introspection, I also sense sadness, and self-contempt.

"Everyone into the *f*MRI!": Delicacy of Taste and Delicacy of Empathy

Now everyone climb back into the *f*MRI as the philosopher's cerebroscope comes into its own in a thought-experiment. There are at least three relevant scenarios. If we both were

in *f*MRI machines viewing the tantrum, the chart of my "read out" would light up areas corresponding to the sadness and frustration and self-contempt sets of neurons whereas yours would remain dark in those areas and register only the anger—you literally did not perceive anything else in that moment. This is not the only possibility. In the case of delicacy of empathy, it is not that one individual makes a better or more educated guess. We are both witnessing an individual throw a temper tantrum. We both obviously perceive the stamping feet, hear the terms of abuse, see the display of anger. It is further hypothesized that this obviously angry individual is masking an underlying sadness and self-contempt by the overt display. Is the one individual able to sense a fine distinction that is literally not available to the other? (1) If we both were both viewing the rage, the prediction is that my "read out" would light up brain areas corresponding both to the rage and sadness, etc. whereas yours would remain dark in those areas and register only the rage—you literally did not perceive anything else in that moment. This is not the only possibility. (2) Alternatively, if the relevant affective areas did light up in you and you sincerely reported that you just saw an enraged guy, not an enraged guy also experiencing sadness, etc., then we would have warrant to conclude that your delicacy of empathy was on the margin of proximal development and with training—no guarantees here—you might be able to raise these pre-conscious impressions beyond the threshold of awareness through the further practice of introspective self-monitoring. The prediction is that you were receptive to something = x that did in fact cause mirror neurons to discharge, but the experience was not yet available to your introspective attention. But with practice the latter would be able to be improved up to some empirical threshold determined by experience. (3) Finally, if the relevant affective areas did not light up and you sincerely reported that you saw an enraged guy masking sadness, etc., then we would have warrant to conclude you were making an educated guess based on inference from behavior clues (not vicarious experience) or perhaps even your own past experience. Your mirror neurons did not discharge, but you cognitive processes enabled you to make an immediate inference of the understanding to an underlying, hidden affective state, immediately masked by the rage. Of course, in the latter case, other areas of the brain would light up corresponding to the executive function of inference making or autobiographical memory.

If I am expert user of empathy, then it is likely that I will directly perceive sadness and self-contempt in the other. However, if I am not an expert, then I will probably have to monitor my vicarious experience attentionally to capture a trace of sadness and self-contempt available to my introspective awareness for further cognitive processing and not otherwise attributable to myself. In this scenario, I have a vicarious experience of sadness and self-contempt that I can access through introspective attention and articulate an interpretation, "That is a sad, frustrated person with self-esteem issues" whereas you just assert he is enraged.

It is a matter of empirical research the extent to which further training in introspective focusing can enhance this attention to detail—this delicacy of empathy. It is likely that training and practice can improve the result, though it is equally obvious that there is a threshold beyond which training will not make a difference. Finally, there is no easy answer to someone's deception when the human is sincerely fooling himself. There is also no easy unmasking of deception when the other is trained to employ the Stanislavsky Method of acting to cause a smile based on genuine recollections of happy times in the past. Interpretation remains an essential part of the empathic process and the indeterminacy of translation is inevitable when one requires that the cause of the emotion be a part of the interpretive context. More on this indeterminacy below.

One exception, which proves the rule, that neuronal discharge is not directly available to awareness, is provided by one of the visual artifacts of migraine headaches. Intense migraines are accompanied by visual phenomena. If a further motivation is needed that brain processes percolate up ("resonate") out of the depths of the underlying infrastructure and that these echoes of form are accessible to introspection, admittedly under highly specific conditions, then consider the fortification images that accompany some migraine headaches. The visible structure is literally forced on the patient in the form of a migraine headache:

> In these headaches, when the visual areas become active, the individual sees a spectacular array of scintillating bars and corners; the display as a whole has the appearance of pre-20th-century fortifications seen from above….Hubert Airy related the displays to known structure of the brain and concluded by recognizing that the spontaneous perceptions probably represented a realistic "photograph" of processing in the brain.[19]

Here the work is to interpret what is available in perception as reflecting the structure of the nervous system. Such images are "lit up" through the self-stimulation of the nervous system as waves of electrical activity advance across the visual cortex that is itself a honeycomb, hexagonal lattice structure.

This of course goes against the conventional wisdom that the one thing the mind is *not* is an inner theatre. However, it does not require a ghost or homunculus, resulting in an infinite regress of executive functionality, since the action is unidirectional, from the underlying physical process and structure to the introspective attention that identifies the content as it enters attentional awareness. The body itself become the screen, employing the eyelids and the self-stimulation of the retina is a visual illusion, if not the theatre, upon which are projected ("represented") the processes of consciousness.

Ekman's Facial Micro Expressions as Input to Empathy

A further area of neurological research that rehabilitates the use of empathy is in the processing of affective and sensory micro expressions. While individuals have control over a wide range of facial expressions and this range can be extended with practice, some, perhaps many, muscles in the face are independent of an individual's voluntary control. These muscles are "hard-wired" to the nervous system and form micro expressions that take less than 0.2 seconds to discharge. For examples, the some of the muscles around the eye are only activated by a smile that recruits real enjoyment as opposed to turning one's lips up into a "grin and bear it expression" as reported by Paul Ekman, whose research broke new ground around micro expressions:

[19] W. Richards. (1971). "The Fortification Illusions of Migraines" in *Perception: Mechanisms and Models: Readings From Scientific American*, ed. R. Held & W. Richards, San Francisco: W.H. Freeman, 1971: 195-202.

> Our research confirmed Duchenne's assertion that no one can voluntarily contract the *orbicularis oculi* muscle (it "does not obey the will"), although it is only part of that muscle that it hard to contract voluntarily.[20]

A similar consideration applies to the red margin of the lips that become narrower in anger as the lips become thinner, which is a very hard action to inhibit. The rate of blinking increases when an individual is nervous. Eyes widen with excitement. Nostrils flare with excitement. Such micro expressions are fleeting and highly transitory. In everyday situations they can reveal a person's underlying emotion, ambivalence, or mixed emotions. For instance, the would-be terrorist has nothing but contempt for those he is about to blow up and a brief expression of it flickers across his face. This directly connects the discharge of neuronal activity – presumably those would be activated in the law enforcement official screening the secret bomber if he were sufficiently sensitive – with the downstream effects visible to the observer who is suitably astute to the fine-grained details.

In another example, a video was made of the proceeding, enabling an engaging after the fact analysis. An individual contemplating suicide but feigning happiness in order to get a weekend pass home from the psychiatric ward was video taped as part of an unrelated training exercise:

> …[I]nspecting each gesture and expression in slow-motion to uncover any possible clues to deceit. In a moment's pause before replying to her doctor's question about her plans for the future, we saw in slow-motion a fleeting facial expression of despair, so quick that we had missed seeing it the first few times we examined the film. Once we had the idea that concealed feelings might be evident in these very brief micro expressions, we searched and found many more, typically covered in an instant by a smile.[21]

---

[20] P. Ekman. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*, New York: Henry Holt, 2003: 206. [Editorial note: This example is also quoted in Chapter ___ Empathy and Expression: Unexpressed Emotions are Incomplete.]

[21] P. Ekman. (1985). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, New York, W.W. Norton, 1985: 17. The phenomenon of micro expressions have broken into the popular press with Malcolm Gladwell's *Blink*.

Prior to this review, she confessed she was still desperately unhappy and had wanted to get out so she could commit suicide. This promoted the review of the training material. In this case, the tape was literally slowed down in order to reveal what was present but did not reach the threshold of conscious awareness. In some instances, this explains the "hunch," "intuition," or "gut feeling" that causes the medical personnel or law enforcement official to stop an individual and ask a few more innocent questions. One viable description is that the trace of the experience of sadness or contempt would be in the system of the perceiving individual and available for further processing by the executive function but would be on the borderline of awareness where it is available for introspective attention if one decides to look to see what is there and if what is there is sufficiently coarse-grained to be experienced. As noted above, it is a matter of empirical research the extent to which further training in attentional focusing can enhance this attention to detail—this delicacy of empathy.

As a venture into experimental philosophy, it is a further thought experiment that an *f*MRI readout would distinguish the perception of a mirco expression of contempt = x amidst the mask of happiness = y. This would be so even if the individual's capture of the micro expression did not reach the threshold of conscious awareness. With practice an individual can perceive a more-and-more detailed time slice. Granted that the human eye cannot interpret more than about 16 frames per second, if you slow down the film, the viewer can penetrate up to an order of magnitude in the direction of fine-grained micro expressions. This is a reversal of our contingency. It motivates ("explains") the benefit of studying art and music for increasing our empathy with others, since our sensory and affective discriminations in the realm of humanlike artifacts—works of art—are readily transferable to the domain of human interrelatedness. This is not an either or proposition—one can spend the morning at the *f*MRI lab sticking one's head in a giant magnet and the afternoon at the Art Institute refining one perceptions (as well as recovering from the morning).

Micro expressions include vicarious experiences of what the other is feeling as available for introspective awareness through monitoring of an individual's experience. These

inputs—vicarious experiences—are available for further processing, including cognitive processing. These tell us that the other is (e.g.) experiencing anger, but it does not tell us why. It is an additional and fundamental input to the process of understanding the other, but I need to look to other inputs, factors, circumstances in order to understand why the other individual is feeling angry. This includes what the other says, what the other shows, as well as the context and the consideration of the relations, whether consistent or not, between these variables.

Putting Sharing into the Shared Manifold

What is the outcome of this meander through neurology? As the neurological infrastructure—computations, processes, and structure—have become better understood, the role of introspection has been substantially reduced from the early days of Locke, Hume, Brentano, etc. when it was supposed to provide access "all the way down" to what was occurring. Early psychologists who tried to use introspection to disclose the micro structure of cognition over-estimated the powers of introspection in a significant way. However, as introspection has been limited by neurology, confabulation, and the invalidation of the homunculus ghost in the machine, it has gained new credibility as providing a distinct form of valid input to the process of empathizing with others. By setting a limit to introspection, these neurological results have made it possible to think both sides of the boundary between the physical and the interpersonal, thus opening up a logical space, albeit a limited one, for the operation of introspection of vicarious experience in which the other is disclosed as other.

Micro expressions are hypothesized to occasion and indeed to cause vicarious experiences of what the other is feeling. These vicarious sensations, affects, and experiences are available for introspective awareness through monitoring of an individual's experience. These inputs—vicarious experiences—are available for further processing, including cognitive processing. These tell an individual that the other is (e.g.) experiencing anger, but they do not tell the individual why. It is an additional and fundamental input to the process of understanding the other, but I need to look to other

inputs, factors, circumstances in order to understand why the other individual is feeling (e.g.) angry. This may include what the other tells me—I am angry because my father doesn't love me; I am angry because infidels are occupying the holy lands; I am angry because Mom liked by brother best; and I may compare this assertion with the context, with what is said as well as with what is shown. Nope, that is not the reason; it must be something else that we have not yet grasped.

Redescriptions in terms of folk psychology loom large here. Two individuals who are having a conversation will negotiate about what is the canonical description of what happened. "Mom gave you the bigger piece of cake – that *means* she likes you best." "No, it doesn't. It means she wants to avoid the guilt trip you always try to run on her. She just doesn't want the confrontation."

What is so far missing is any mechanism intermediate between folk psychology and the underlying hardware of the brain, which is central to the activation and implementation of empathy. Presumably the strongest candidate mechanism is the shared manifold. [22] The mechanism that is supposed to help the one and the other articulate a common interpretation and converge on one based on the shared manifold. This is a theoretic construct that is supposed to underlay the other aspects of the functional architecture of empathy such as self-awareness, flexibility, and emotional control.

From the perspective of the phenomenal experience of empathy (not to be confused with the phenomenological method of Husserl), Gallese proceeds to identify the shared manifold with an enlarged sense of empathy (Gallese 2001: 45). This is definitely *not* one of the consequences of neurology for empathy. If anything, it is a bold standalone proposal--to enlarge the application of empathy to provide a foundation for community relations in the sense of intersubjectivity at large. This proposal has merit entirely independent of mirror neurons and may present transcendental challenges that are unattractive to hardnosed scientists. Unfortunately, by identifying empathy with the

[22] V. Gallese. (2001). "The 'Shared Manifold' Hypothesis: From Mirror Neurons to Empathy*," Journal of Consciousness Studies* 8, No. 5-7, (2001) 30-50.

shared manifold (as it will subsequently be unpacked) the risk is that empathy will be reduced to various of its precursors and derivatives such as vicarious experience, emotional contagion, communicability of affect in art, or ethically relevant sentiments such as respect or fellow-felling. However, the enlargement—this is not a dirty joke—looks especially motivated if one looks at the contribution of psychopathology and psychiatry in regarding autism, antisocial behavior ("sociopathy"), and the ability to make sense of sensory manifolds such as the human face as failures of one or another components of empathy. Note it is not necessary to regard empathy as a single module that is missing or broken, nor is such a scenario likely to be the case, given the variety of symptoms.[23]

In folk (social) psychological terms, a vicarious experience of what the other is experiencing is projected onto or grounded in the shared manifold. On the upper end the shared manifold itself becomes input to further interpretive activities that are grouped as "cognitive empathy," putting oneself in the other's shoes, answering how I would feel if the events you are experiencing also happened to me, answering how you feel about it, and other imaginative variations (these are not mutually substitutable one for another). This is a theoretic construct that is best deployed in supporting other aspects of the functional architecture of empathy such as self-awareness, flexibility, and emotional control (Decety and Jackson 2004: 71f.) and can be extended to the infrastructure of folk (social) psychological concepts such as emotional contagion, animal magnetism, fellow feeling, and vicarious experience. (This list is not complete).

In his first formulation of the shared manifold, Gallese invokes the works of Husserl, Edith Stein, George Herbert Meade, and others in describing different forms of empathy—that which is to be explained.

[23] For an argument that empathy is the foundation of intersubjectivity in the sense of interrelations between humans in a community—a hermeneutic rather than a social-neuro-scientific approach see, L. Agosta. (1984). "Empathy and intersubjectivity" in *Empathy I*, ed. J. Lichtenberg et al. Hillsdale, NJ: Lawrence Erlbaum Press, 1984: 43f. Agosta does not explicitly invoke empathy in the "enlarged way" proposed by Gallese, since empathy has always been "enlarged" enough to ground community.

These thinkers do indeed provide support and inspiration for an approach to intersubjectivity that allows one individual to directly experience the other individual. Unfortunately, the Husserl who wrote *Ideas II* ran into issues around solipsism advancing from an analysis of the lived body—Gallese's embodied simulation--of the epistemological subject to that of the other. In short, even though the rich kinasethetic experience of the subject's own lived body provides a basis for self-consciousness as the individual moves through space, Husserl ran up against the accusation of solipsism, an accusation that had some merit. The accusation had validity. After performing the phenomenological reduction in its various forms to disclose the world as an ideal meaning, the self is isolated; other individuals become problematic both ontologically and epistemologically. Have I not accidentally excluded these other individual centers of awareness, agency, spontaneity, meaning-giving mental acts, in so far as they are part of the natural world that was just bracketed? By referring to one of the more obscure and incomplete attempts to found intersubjectivity, *Ideas II*, Gallese's philosophy colleagues may have been playing a little philosophical joke on him by locating the shared manifold in a work that was not widely engaged due to accusations of solipsism.

However, even if it is impossible to build on *Ideas II*, Husserl addressed the issue in a separate treatise *The Cartesian Mediations* as well as three additional volumes in his unpublished *Nachlass* in which *Einfühlung* plays a central role.[24] Aside from the issue of tightening up the bibliographical reference, Gallese is on firmer group with Edith Stein and Merleu-Ponty. The lived body means that I do not just see upturned edges of the other individual's mouth, narrowed but open eyes, and relaxed eye brows, I see a smile. I see that the person is happy to see me. I do not just see a lever with attached two pronged-grabbing extensions raising a box with the fulcrum at the point of the hips; rather I see a workman straining to lift a heavy box of books. Finally, with Meade the problem of other minds does not occur at all. The other is a part of the self system from the start. Not only is the individual a part of the group (community); but the group is a constituent of the individual. For Mead, it is necessary for one individual to have

[24] E. Husserl, *Zur Phänomenologie der Intersubjectivität: Texte aus dem Naclass: Erseter Teil: 1905-1920; Zweiter Teil: 1921-1928; Dritter Teil: 1929-1935*. ed. I. Kern. The Hague: Martinus Nijhoff, 1973.

available a representation of the other in one's own self in order for gestures to be significant symbols. This representation is the "generalized other":

> From the behavioristic standpoint it [i.e., signification] must take place through the individual generalizing himself in his attitude of the other . . . through taking the attitude of what may be called the generalized other. . . . He finds himself speaking to himself and to others with the authority of the group . . .[25]

By means of this model of the other the social process is brought within the experience of the individual. The other's response is made a constituent of my own experience. If I am to participate in significant communication, then it must be a dialogue not a soliloquy. The self and the other are differentiated out of a common social process of interrelation. If anything, priority belongs to the other—we tend to live for the most part in the expectations and demands that others make on us, and only later become aware of ourselves as individuals.

The shared manifold deserves to be assessed on its own merits. As described by Gallese (and to a lesser extent Decety), the manifold is an activation of mirror neurons that are isomorphic in function to those in a perceived source, another individual performing an action. In the case of motor neurons, processing does not pass the threshold available to introspective awareness. I see you grab for a cup to drink, and my mirror neuron are activated. Though I am obviously unaware of the discharge of neurons, I immediately and intuitively am aware of what it *means* that you are grabbing a cup. You *want* to drink or *want* to show me a nice piece of porcelain or even *want* to demonstrate hand-eye coordination. According to Gallese, my understanding of what you are doing is an "embodied simulation," not an argument by analogy (Gallese 2007: 459). I immediately appreciate you are an "embodied" individual just as I am without the requirement for any analogical inference from my experience to yours. A similar consideration applies in the case of observing you experience a painful stimulus, for example, in slamming a car door

[25] G.H. Mead. (1922). "A Behavioristic account of the significant symbol," *Journal of Philosophy* 19 (1922), 157-63: 161-62. Mead also identifies the child's imaginary friend as evidence that an individual always carry the other around with him and eventually learns not to talk aloud to someone who is not (physically) present; however, the conversation goes on. See G.H. Mead. (1912). "The mechanism of social consciousness*," Journal of Philosophy, Psychology, and Scientific Method* 9 (1912): 403.

on your finger. Ouch! The finger give a vicarious throb just thinking about it, albeit at a less intense degree than the actual victim. A source and a receiver are related. The source causes the receiver to mirror the pattern of activation. In this way, the source becomes the target of empathy as further processing of the experience occurs in terms of the self-other distinction, interpretation of the manifold as the possibility of human interrelatedness, and disclosure that the other is experiencing something of which I have a vicarious experience.

In order to explain my experiencing the other's pain vicariously, the shared manifold is functionally constellated in my own system of mirror neurons, resonating with the other human being's pain, being receptive to the input provided "from below" by the neurological infrastructure in the (empathizing) human being whose mirror neurons are activated. The human being that is the source doing the slamming on his own finger is the cause of the shared manifold and functionally participates precisely because he is the cause. However, he does not phenomenally participate—at least not yet. He is aware he is hurting; but he is not aware that anyone else is aware. We have seen numerous examples of the communication of a sensation or affect without introspective attention – inner imitation, micro expressions, fortification illusions in migraines. An individual can be the target of empathy without being aware of it.

The manifold is bounded by the mirror neurons of the perceiving subject where, in folk-psychological terms, it gives rise to a vicarious experience of what the other is (phenomenally) experiencing. As in the above-cited example, if I look on helplessly as my neighbor slams the car door on his finger, I wince and get a vicarious pinch in my finger, granted that it is much less intense than the one he is now experiencing. Where is the manifold located? Gallese states that "mirror neurons instantiate a multimodal intentional shared space...analogous neural networks are at work to generate multimodal emotional and sensitive shared spaces" (Gallese 2007: 462).

In order to enable my experiencing the other's pain, the shared manifold is constellated in my own system of mirror neurons, which, in turn, resonate with his pain, providing the

neurological infrastructure in the individual whose mirror neurons are activated. Unfortunately, that means it is not shared, since the one who is taking the action—doing the slamming in this case--is the cause of the manifold but does not participate in it. The one with whom the empathizing is occurring falls out of the equation, which, once again, is not fair to the individual who is supposed to benefit from the empathy. That is not an encouraging result in terms of establishing a community of subjects. As an explanatory mechanism, more work is required on the shared manifold.

Thus, Jean Decety, who builds on and extends the work of Gallese using *f*MRI technology, is spot on to make emotional regulation ("control") an essential part of the definition of empathy (Decety & Jackson 2004: 71). The assertion is that the empathy requires that the empathizer has emotional control over his or her feelings and affects in order to avoid empathic distress. The key term here is "control." Too much intensity to the affective communication and the result is "empathic distress," in which the subject is overwhelmed by the experience of the other. For example, in an extreme case going beyond a hurt finger, I am so distressed by the other's loss I am literally unable to be with the other's experience (e.g.) of the death of her or his father. Too little intensity, and the result is apathy, in which the subject is indifferent, unmoved, untouched by the experience of the other. In either case, the result is simply leaving the other to his own devices, abandoning the scene or even blaming the victim. It is also the basis of blind spots in the psychology of everyday life and counter-transference in psychoanalysis.

Thus, the view of Decety *et al.* has become more cautious since publishing "The Functional Architecture of Human Empathy" in 2004.

> …[N]onrepresentation…representational and metarepresentational mechanisms…need to be taken into account for a full understanding of human empathy. Representations, in this context, are defined as parallel distributed patterns of activation that reliably fire in response to a given stimulus. These neural networks encode information and, when temporarily activated, enable access to this stored information. This distributed nature of social processing poses a challenge to understanding the neuroscience of social cognition in general and empathy in particular. In addition, social psychological concepts, such as empathy, do not necessarily correspond to neuronal processes. Relations between

> psychological and biological processes cannot be understood fully by investigation at a single level of organization. Therefore, multilevel research is necessary to form bridges among disciplines and ultimately achieve a truly interdisciplinary social neuroscience.[26]

Amen to that! In the 2004 article each of the components of empathy as defined by Decety and Jackson—shared representation, the self-other distinction, point of view (mental flexibility), and emotional regulation ("control")—are correlated with more-or-less specific neural parts of the brain. In 2004, each of the components of empathy—the shared manifold, self/other distinction, first- and third-person flexibility, and emotional regulation—had specific neural correlates of social cognition, even if the executive functions of regulation were still coarse-grained.

Fast forward to 2006: "empathy…[does] not necessarily correspond to neuronal processes." At no point is it suggested that empathy can be reduced to these neural correlates of empathic awareness; and the discourse remains at the level of statistical generalizations. Still evidence from neuropathology and psychiatry suggest that brains states are causing empathy, at least as necessary conditions, and that without the underlying computations of the different parts of the neural network, the empathy would not appear or would be limited at least at the level of statistical generalization. Singular causal events are occurring in both directions.

I do not want the level of granularity to become an obstacle to teasing out the significance. The history of advances in neurology provides strong evidence that an ever widening array of mental functions are being driven down to the level of the hardware in neural correlates of consciousness, primarily in terms of necessary conditions of functionality.

There are at least two extreme cases here. In the one case, no matter how global the mental functionality—perceiving, imagining, intending—the removal of that particular

---

[26] J. Decety & C. Lamm. (2006). "Human empathy through the lens of social neuroscience," *The ScientificWorld Journal* 6 (2006), 1146-1163: 1148f.

neuron or set of neurons will disable the function. That hypothesis—that every function of consciousness—has a neural correlate is proving to be an extremely fruitful line of research, though one can imagine a lowest level of granularity beneath which organic-functional-representational distinctions collapse or imagine reaching some principle of indeterminacy. That leads to the other extreme—social functions and entities at the same level as folk psychology that would lack any one-to-one correlation. Individual functions would emerge holistically from varying arrays of networks such that they are indistinguishable at the micro-level. The cerebroscope would provide a read out of neural activity that did not translate all the details to the lower level. Even if the level of granularity is as excruciatingly detailed as one can imagine such that "all," "some," and "none" are implemented by specific neurons, then the pattern or mirror neurons shows that you are experiencing a qualitatively similar experience to what he is experiencing. It does not show the equivalent redescriptions of the meaning of the propositional content.

What is missing – other than the call for further interdisciplinary exploration, which covers anything that might under any imaginable interpretation be missing - is empathy as an enabler and implementer of what is human. In its context of human interrelatedness, empathy is a committed listening that discloses the other in the affect of respect to understand the other as possibility and let the other be in equilibrium as an interpretive choosing of authentic selfhood in the face of commitment. Empathic listening is what restores the other to equilibrium. This is just a different approach and definition of empathy that does not contradict the functional architecture, but is at a different level of organization and, I suggest, orthogonal to the functional architecture. It captures the aspects of empathy that are a human process—listening, a paradigm affect that clears away other specific feelings, restores emotional equilibrium, and allows whatever feelings are significant to show up and get unpacked and implemented as possibilities that provide for choice and commitment. Maybe there is a way to operationalize such a process-oriented definition. Once again an *f*MRI machine reduced to the size of a helmet so two people can interact would be useful.

Even if you can translate all of the interactions into a cerebroscope readout, there are dynamic properties—what human beings mean by the relationship between self and other—that can be represented as a numeric readout, but cannot be determinately and disambiguously translated into human terms without using the very concepts such as empathy, respect, understanding, that were supposed to be explained by the readout. That is necessarily so even if a *f*MRI could be miniaturized as a helmet and worn by the two people in dialogical interaction.

This does not merely say that the number of redescriptions is potentially infinite or that the consequences of a given description are infinite and escape, though it says that too. This does not say that "peak experiences" such as ecstasy, deep meditative enlightenment, or love are unable to be captured, since such experiences would indeed show up as a spike in the read out. This does not say that the allegedly complete list of numeric readouts, including the potentially infinite packets that could be added to the list, would be insufficient since the scientist could diagonalize across the list, demonstrating that it was incomplete. That is an issue but there is an additional problem. The issue is that there is no necessary lawlike pattern between the interaction of social functions and components at the level of folk-psychology and that at the level of the brain (and any relevant associated physiological components). Thus:

> Theoretically, we propose that the meaning of a given object, action, or social situation may be, to some extent, shared by several individuals and thus should activate the same neural network in their respective brains.[27]

This requires a bit of work to disentangle the position. That meaning is shared is a common place. There must be more going on here. This is not a statement about a general, lawlike causal relationship between meaning and a brain process. There is still only correlation, possibly between singular mental events that cause one another, each of which has a neurological correlate that itself may (or may not) have a general lawlike

[27] J. Decety and T. Chaminade. (2003). "When the self represents the other: A new cognitive neuroscience view on psychological identification," *Consciousness and Cognition* 12 (2003) 577-596: 581).

causal relationship that operates horizontally at the neurological level; and that correlation is between a shared representation, a meaning, and the brain's neural network.

The shared manifold proposes to provide a bridge between the vicarious experience of the empathizing individual and the other who is the target of empathy. As the vicarious experience is interpreted and communicated to the other—the target of the empathy—it enables the other to use the shared manifold to restore his or her emotional equilibrium when it has been temporarily knocked out of homeostasis by the slings and arrows of outrageous fortune. The shared manifold, however, is an opaque (even messy) structure that resembles a causal thicket with trial and error and statistical generalization rather than rule-governed, general laws. The shared manifold activates an intermediate level of function between neurology and folk social functioning ("sociology") that is "the same mental code" (Decety, 2003: 584 echoing the work of W. Prinz) or an instance of a "multimodal intentional shared space" of "analogous neural networks…[that] generate multimodal emotional and sensitive shared spaces" (Gallese 2007: 462). Here we cross levels of organization from the relatively well-defined domains of neurology and social distinctions appropriate to human social interrelations at the level of folk (social) psychology into that of the shared manifold. This is not to say that shared manifold is not a useful hypothesis; but it is also another name for our ignorance.

In Search of a Bigger Magnet

After elaborating a complex functional mechanism of significant power – the shared manifold – Gallese concludes that empathy does not rely on inference nor, in particular, on an argument by analogy (Gallese 2007: 459); but empathy provides the immediate givenness of the other in his affect, emotion, or feeling ("embodied simulation"). The *non sequitur* is that the former is at the level of our phenomenal experience whereas the latter is at the subpersonal level of mirror neurons. I do not experience an inference – the embodied simulation lacks an phenomenology other than seeing you grab the cup to drink (obviously); but, under the hood, significant simulations are being computed and they have the form of an isomorphic activiation of parallel neurons, in short, an analogy.

Analogies and inference – not the same thing but closely related – are occurring in abundance. An example will be useful.

In a crucial passage in Thomas Mann's Buddenbrooks (1901), Hanno Buddenbrooks experiences empathy with his father (also named "Thomas"). The father and son have been at odds for years. Hanno's artistic temperament mirrors that of his father; but the father has renounced and denied his artistic aspirations in the interest of running the family business. His one concession was to marry the beautiful and artistic Gerda Aronsen. However, things are not going well, and Gerda, even if not having an affair, has crossed the line of social impropriety, playing passionate duets on the violin with her friend, the Lieutenant. We join the text:

> But one thing they both felt: in the long second when their eyes met, all constraint, coldness, and misunderstanding melted away. Hanno might fail his father in all that demanded vitality, energy and strength. But where fear and suffering were in question, there Thomas Buddenbrook could count on the trust and devotion [*des Vertrauen und Hingabe*] of his son. On that common ground they met as one" (Thomas Mann. (1901). *Buddenbrooks*, tr. H. T. Lowe-Porter, New York: Random House, 1961: 507; slightly modified to include *Vertrauen* [trust], which was omitted from the translation).

One might argue that "under the hood" his human biocomputer and its mirror neurons are simulating the distress in the voice and face of his father, enabling Hanno to "get it." And, no doubt, mechanistic processes are being discharged. If Hanno and his father were able to plug miniature functional magnetic resonance imaging machines (say) into their ears and the read outs were captured, then the moment at which their eyes touched would indeed show a spike in the discharge of mirror neurons. Something happened and it is represented in this gigabyte of data. What it would *not* show is the human meaning as "all constraint, coldness, and misunderstanding melted away." What it would not show is empathic understanding of the possibility that this big, strong guy suffers too. What it would not show is the possibility of a human relationship between father and son as vulnerable individuals that trust one another. There are many reasons for this. One is that the *f*MRI has the temporal constraint character of a tool being used right now; whereas Hanno and Thomas are open to time across all three temporal dimensions which, in turn,

enables them to put the past back into the past and create a new possibility, living into the future, a loving and affectionate relationship instead of a hostile and fearful one.[28]

In the same vein, Decety advocates social neuroscience since he rejects (or would reject) "nothing but-ism," the reduction of social relations to nothing but the interaction of individual organisms. This is well and good. But it simply does not follow that concepts such as empathy, affectivity, or intentionality are explained in any lawlike way by a shared manifold ("a multimodal intentional shared space" (Gallese 2007: 462)).

This points to a different approach and use of empathy that does not contradict the functional architecture of Decety, but is at a different level of organization and, I suggest, orthogonal to the functional architecture. It captures the aspects of empathy that are a human process. Relating to the other individual as to the possibility of spontaneously beginning something new; having respect for the other as a paradigm affect that clears away other specific feelings, restores affective-sensory equilibrium; and is implemented as a human possibility that provides for choice and commitment. In its context of human interrelatedness, empathy is a committed listening that discloses the other in the affect of respect; discloses the other as possibility that lets the other be in equilibrium as an interpretive choosing of authentic individuality in the face of commitment. Empathic listening is what restores the other to equilibrium. Maybe there is a way to operationalize such a process-oriented definition. Once again a powerful *f*MRI machine scaled up to the size of an entire room so that two people can interact in it would be useful. But there is a "but."

As noted above, here we cross levels of organization from the relatively well-defined domains of neurology and social distinctions appropriate to human social interrelations at the level of folk (social) psychology into that of the shared manifold. This is a dense, interactive "causal thicket" of several interrelated domains -- social interrelations, our phenomenal experience of the other, the emergent properties of being human, vicarious

[28] Editorial note: This material is quoted in Chapter ___ on the Empathy and Intentionality. One occurrence should be deleted and a cross reference substituted. This is a long passage and detailed argument, and there may be value in embedding the complete section in Chapter ___. The entire passage is about 1900 words.

experience, intentionality, and the meaning, including linguistic meaning, with which we humans give an account of our experience.[29]

Even if you can translate all of the interactions into a *f*MRI ("cerebroscope") readout, there are emergent properties—what human beings make the relationship between individual and other *mean* in human terms—that can indeed be represented as a numeric readout, etc., but cannot be determinately and disambiguously translated from human terms without using the very concepts such as empathy, respect, vicarious experience, human interrelatedness, etc., that are supposed to be explained by the readout. We are spared disentangling the rich semantic impertinence of the shared manifold. This is because the interposition of a shared manifold as a hybrid structure participating in folk (social) psychology and neurology will not solve the problem of explaining empathy through neuroscience and will itself be engulfed by it. Consider an example.

If the empathic glimmer in the eye of the other makes my heart skip a beat, the neurological equivalent of that interpretation—"skip a beat"—will show up on the *f*MRI. Here "skip a beat" will be used as an arbitrary sign for whatever physiological complex is being interoceptively perceived rather than as a poetic metaphor of romantic infatuation. What will be missed, however, in the translation from human interrelatedness—e.g., seeing the other's happiness-- is precisely the human causal connection between the glimmer of the one human being and the heart beat of the other. She will eye me, I will catch my breath, and the data storage will be complete. No data will be lost. It will just not capture the human element. Let's be clear on this. This is *not* just a function of the contingent circumstance that we have only one, limited *f*MRI machine. Nor is this just a matter of eliminating metaphorical language. This does *not* merely say that the number of redescriptions of a human relation is potentially infinite or that the consequences of a given description are infinite and escape, though it says that too.

[29] "Causal thicket" is a term by W. Wimsatt, "The Ontology of Complex Systems," in *Reengineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press, 2007: 200.

Within the research program and its hypothesis of materialist, nonreductive correlation, with which I agree, all of the social components - empathy, etc.- will have neural correlates of consciousness (NCCs) without exception and in sequence. This does *not* say that "peak experiences" such as ecstasy, deep meditative enlightenment, or love are unable to be captured. Such experiences would indeed light up the *f*MRI like a Christmas tree. There is an additional problem.

As noted previously, one way of posing the problem is to say that there is no necessary lawlike pattern between the interaction of social functions such as empathy at the level of folk (social) psychology (or even at the intermediate functional level of the shared manifold) and the NCCs at the level of neurology. In comparison with the application of empathy to human interrelatedness, the neurological read out indicating what areas light up will not be informative in the terms required by human relations. For example, the *f*MRI read out will provide explicit data the upturned edges of the other individual's mouth, narrowed but open eyes, dilated pupils, and relaxed eye brows, in even more detail, whereas I require information about a human smile. It will not be informative no matter how coarse or fine-grained the slice of neurology, even it were down to a single mirror neuron (though that is unlikely given the high "fan out" out of neural networks).

This has further philosophical implications. Under one interpretation that goes beyond Gallese or Decety, but, arguably, is consistent with their overall research program(s), empathy is a serial processor on top of a parallel distributed processor. Empathy is the capacity in which the parallel neural network processing system of our sensory-motor affective systems intersects with the serial processing system of formal cognition. The neural network is responsible for coordinating inputs for all our sensory modalities including how we feel about our interactions with others in our environment. What gets semantically interpreted by empathy will be patterns of activation of such units as facial expression, tone of voice, propositional content, bodily posture, aspects of behavior as integrated in context. In a computational reinterpretation of the associative tradition (e.g., David Hume), these different features are encoded as the neural network is in effect trained by the good, bad, and indifferent developmental experiences of growing up. This

includes the likelihood that the neural network itself is not a blank slate. The basic emotion of anger will get encoded in the neural network with sometime arbitrary and sometimes calculated propositional contents—explicit semantics based on experience—in a nontrivial way that forms a vast array of patterns, perhaps finite but still combinatorially intractable—and based on the historical trajectory of the human being through its developmental history. The emotion – e.g., anger – will emerge from a continuum of activation patterns corresponding to a wide variety of sensory modalities. This output will itself provide input to the serial processor. As a serial processor, empathy will apply perspective taking (first-person, second-person, etc.), delay of gratification and desire, and rules of logic to the inputs provided by the parallel processor.

Thus, the folk (social) psychological layer—including such mixed up concepts as empathy, affectivity, smiling--need not correspond to any particular projection or representation in the underlying neurology. The neurology is nevertheless a necessary condition of the folk (social) psychological phenomena of empathy. But empathy is not directly projectable onto the neural network whose interrelated activation constitutes the expression, representation, and capture of the emotion.[30]

The shared manifold casts some long shadows: downward in the direction of mirror neurons and the projection of interoceptive experiences onto a map of the body ("embodied simulation"); and upward in the direction of a bridge between the first-, second-, and third-person perspectives, awareness of the self versus the other, and interpersonal mirroring in psychotherapy. The more sharing that is articulated and embedded in the shared manifold as folk (social) psychological distinctions are deployed in it, the less relevant from the perspective of providing a causal explanation it becomes. The more crisp and explanatorily relevant the neurological distinctions as activation of isomorphic mirror neurons, the less anything of human interest is captured or articulated by it. In that sense, the shared manifold is semantically transparent – we see through it without awareness, but causally opaque (i.e., not clear how it works). We would have

[30] A. Clark. (1988). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, MA: The MIT Press, 1989: 49, 50, 70.

hoped for the reverse – semantic opacity so that we might encounter the meaning in awareness and causal transparency so that we do not need to invent additional mechanisms.

The neurology will be governed by general laws of neurology, and the human empathy that corresponds component-by-component to the sets of neurons governed by these laws—as demonstrated by Decety and Jackson (2004: 71f.)--will remain indeterminate. Yes, there will be contingent, statistical generalizations possible between my empathy and your embodied self-expression. These will be useful as rules of thumb and for training purposes. The medical benefits of the research program of social neuroscience will continue to unfold. But every instance of human empathy, like every action and event that is a function of human choice, will be able to be described as spontaneous and intentional while the corresponding neural events are involuntary and unintentional. Our search is disappointed. A bigger magnet will not help.