Interpersonal comparisons of well-being, the evaluative attitudes, and type correspondence between mind and brain

JP Sevilla*

This draft: December 10, 2009

Number of words in this document: 11,273

---

*Research Associate, Department of Global Health and Population, Harvard School of Public Health

0. Summary

Interpersonal comparisons of well-being (ICWs) confront the longstanding unsolved *epistemic problem of other minds* (EPOM): the problem of how to achieve objective knowledge of people's subjective mental states. The intractability of the EPOM may lead to the hope that *Rational Choice Theory* (RCT) can show that information about how people would choose over goods and gambles is sufficient--and information about subjective mental states therefore unnecessary--for interpersonal comparisons of levels and changes in well-being, thereby bypassing the EPOM. I argue that this hope cannot be fulfilled. Our most plausible theories of value--whether anti-realist or realist--and theories of what makes a life go best--whether preference hedonism, success theory, or objective list theory--tie well-being to our *evaluative attitudes* towards our lives. These are distinct from and only contingently related to *motivational attitudes* to choose or behave in certain ways and therefore to choices and behaviors themselves. Interpersonal comparisons of the evaluative attitudes are therefore necessary, though perhaps insufficient, for ICWs. Preference theory's *zero-one rule* ignores these attitudes and is therefore implausible. Its *extended preference* approach assumes that our preferences are perfectly sympathetic and therefore begs the question of the EPOM. I argue that a principled solution to the EPOM, and to interpersonal comparisons of the evaluative attitudes, is provided by *type correspondence* between these attitudes and brain states. It remains an open and difficult question whether there exists a summary evaluative attitude whose intensity can serve as an index of an individual's over-all well-being, and which is the appropriate target of all efforts aimed at promoting the personal good, or whether the self and therefore well-being are too fragmented for this.

1. Interpersonal comparisons of well-being and their relevance to social choices

Interpersonal comparisons of well-being (ICWs) are a longstanding unresolved theoretical issue at the center of welfare economics, utilitarianism, and social choice theory. Indeed it is an issue that confronts any theory of distributional or social justice or of the collective social good. Whether ICWs are meaningful is the question of whether it is meaningful to say that the improvement in one person's well-being resulting from some policy would be larger some other person's, or whether one person's level of well-being is lower than some other person's. These involve interpersonal comparisons of changes and levels in well-being respectively. Any reasonable theory of distributional or social justice will give some weight to the goal of promoting the well-being of individuals in society. But the scarcity of resources implies that promoting one person's well-being involves foregoing the opportunity to promote another's. Thus the goal of promoting the well-being of individuals necessarily involves deciding whose well-being should get higher priority than others'. So how should we decide on the priority of different individuals? Utilitarianism says we should prioritize individuals for whom we can produce the largest positive *changes* in well-being. Prioritarianism and leximin say we should give some priority to those at lower *levels* of well-being. Thus setting priorities requires making interpersonal comparisons of both changes and levels of well-being.

2. The epistemic problem of other minds (EPOM)

Up until the early 20th century, the dominant theory of well-being was Benthamite hedonism, according to which well-being equals the net quantity of pleasures and pains within a life. The traditional and central difficulty faced by ICWs, stated as early as Jevons, most influentially by Robbins, and echoed by many since including Hare and

Harsanyi, is what Farah calls the *epistemic problem of other minds* (EPOM)[1]: the problem of whether and how we can have objective knowledge of people's subjective mental states, and particularly for the hedonist how we can objectively compare the magnitudes of people's pleasures and pains. Pleasures, pains, and their intensities seem to be *essentially private*. Each person has access only to his or her own pleasures and pains and not to those of others. At this very moment for example, I cannot tell whether I am happier with my life as a whole, or less happy, or equally happy as the man person who just walked past my table, because I only have access to my happiness but not his. Each of us only has direct experience of our own mental lives, not of those of others. Loosely speaking, the EPOM is the problem of how we can look inside each others' heads, or the problem of our not being mind readers. The EPOM jeopardizes any observer's ability to know what people's lives really look like "from the inside" and therefore that observer's ability to make ICWs. For if ICWs necessarily require interpersonal comparisons of mental states (such as pleasure and pain), and if we do not have direct access to others' mental states, and if behavior is only contingently related to those mental states, then information about behavior is not sufficient to make ICWs. The EPOM remains unsolved to this day, and it is the main reason we do not have a satisfactory philosophical or theoretical basis for a method of performing ICWs.

3. Preference theory and the zero one rule

Belief in the insolubility of the EPOM caused a shift in the utilitarian tradition away from Benthamite hedonism, with the central role it gave to the mental states of pleasures and pains, towards rational choice theory (RCT) which underlies modern economics and which gives the central role to choices. One reason for the shift is clear: people's choices

are publicly accessible in the way that their mental states are not, so that the epistemic problem that afflict the latter do not afflict the former. We may not see whether your pleasures and pains, but we can all see what you order in a restaurant, what you buy from the grocery, what occupation you choose, etc. RCT is a spectacularly successful empirical theory, particularly in predicting economic behaviors such as consumption, savings, and labor supply. It is perhaps the most successful empirical theory in the history of the social sciences. Some economists and philosophers might hope that RCT can also show us a way to ICWs that does not require solving the EPOM but rather bypasses it. The hope is that RCT can show that information about what people would choose is sufficient, and information about their subjective mental states therefore unnecessary, for ICWs. Its best effort at doing so, and therefore the effort upon which this hope must rest, is called the *zero-one rule*[2] which we now describe.

Assume that of all possible goods or life circumstances (I use the words "goods" and "life circumstances" interchangeably), John's best and worst choices are A and C respectively. By this I mean that whenever John is allowed to choose between A and anything else, he would always choose A, and that whenever he is allowed to choose between C and anything else, he would always choose the anything else. If RCT says that information about what people would choose is sufficient to make comparisons of well-being, then it must hold that the fact that John would choose A over C is sufficient to conclude that his well-being at A is higher than his well-being at C. This is an intrapersonal comparison of *levels* of well-being.

What about intrapersonal comparisons of *changes* in well-being? The standard approach is to derive the meaning of intrapersonal comparisons of changes from how

people would choose over gambles. Let us assume that John's choices over gambles satisfy the axioms of expected utility theory. Then for every good G, there is some corresponding unique probability p_G with value between zero and one such that John will be indifferent between G and a gamble with probability p_G of getting his best choice A and a probability $1\text{-}p_G$ of getting his worst choice C. I label the probability with a subscript corresponding to the name of the good to emphasize that the probability is good-specific, so that in general different goods correspond to different probabilities. And the more John prefers a good, the higher the corresponding probability of getting A has to be in the gamble in order for John to be indifferent between the good and the gamble. Since A is John's most preferred good, we have $p_A=1$, and since C is his least preferred good, we have $p_C=0$. Every other good G less preferred than A but more preferred than C will correspond to a $0<p_G<1$. The value p_G corresponding to any good G can be called John's *degree of preference satisfaction* or *utility* or *well-being* with that good G.

John's indifference between G and its corresponding gamble implies that if he were ever to find himself in possession of G, he would accept a $1\text{-}p_G$ risk of the bad outcome that is having G replaced with C, in exchange for a p_G chance of the good outcome that is having G replaced with A. If we make the further assumption that John's makes choices over gamble in order to maximize his expected well-being--an assumption that is not a part of expected utility theory but that Broome[3] has argued to be necessary for intrapersonal comparisons of differences of well-being and calls *Bernouilli's hypothesis*)--then this indifference implies that the product of $1\text{-}p_G$ and the improvement in well-being in moving from C to G is equal in value to the product of p_G and the improvement in

well-being from G to A. This in turn implies that the ratio of the improvement in well-being from G to A over the improvement in well-being from C to G is equal in value to the ratio of $1-p_G$ over p_G. This result regarding ratios of changes in well-being and ratios of probabilities also follows when the place of A in this description is taken by any other good that John prefers to G, and the place of C taken by any other good to which G is preferred. What this paragraph's argument shows is that if we assume that choices over gambles satisfy the assumptions of expected utility and that Bernouilli's hypothesis holds, then ratios of changes in John's well-being are well-defined. This in turn implies that statements about whether some change in John's well-being is larger, or smaller, or equal in size to some other change, is well-defined.

To summarize, RCT shows us how information about an individual's choices over goods are sufficient for making intrapersonal comparisons of levels of well-being, while information about an individual's choices over gambles, when combined with the expected utility axioms and Bernouilli's hypothesis, are sufficient for making intrapersonal comparisons of changes in well-being.

But what can RCT say about *interpersonal* comparisons? To address this, we need to introduce another person: me. Assume that my best and worst choices are B and D respectively, and that like John, my choices over gambles satisfy the expected utility assumptions and Bernouilli's hypothesis. What can RCT say about the comparison between John's well-being at his best choice A and my well-being at my best choice B? As Hausman elaborates, if only information about what people would choose is sufficient to make interpersonal comparisons, and if both John and I are at our respective best choices, then RCT cannot depart from the conclusion that John and I must be equally

well-off. For to consider John better off than me, or vice versa, would require the use of information other than how we would choose, and such extra information is not supposed to be necessary. Thus RCT is compelled to conclude that John at A is equally well-off as I am at B. Analogously, it must conclude that John and I are equally well-off as each other when we are at C and D respectively. If we combine these interpersonal comparisons with the previous intrapersonal comparisons, then RCT must conclude that John is better off at his best choie A than I am at my worst choice D, while I am better off at my best choice B than John is at his worst choice C. It must also conclude that the change in John's well-being in moving from A to C equals that of my moving from B to D. More generally, for goods other than our best and worst, interpersonal comparisons of levels and changes in our well-being can be conducted through comparisons in levels and changes in our degrees of preference satisfaction with our holdings of goods. These are the results of the zero-one rule.

It will help to clarify some points about how we must interpret or understand RCT if it is to successfully bypass the EPOM:

1. RCT can bypass the EPOM *only to the extent* that its basic concepts such as preferences and choices are interpreted in such a manner that they correspond to publicly observable behaviors rather than private or subjective mental states. So preferences can mean no more than empirical descriptions of how individuals would choose when presented with pairs of alternatives. To say that John prefers A to C is to say no more than that he would choose A over C given the opportunity to choose between them. Preferences cannot be interpreted or defined in terms of desires, wishes, predisposition, or wants insofar as these latter terms involve mental states or subjective feelings or first-

person qualitative experiences. For if preferences were defined in terms of these latter terms, then this would imply that preferences and choices are not the same thing, and we would not be able to make claims about the empirical relationship between the two (such as the claim that we tend to choose what we prefer) without looking inside people's heads. I am *not* saying it is unreasonable to interpret preferences in terms of these subjective mental states. I am saying that such arguably reasonable interpretations of preferences would not allow RCT to bypass the EPOM. I shall therefore henceforth eschew such subjectivist interpretations of preferences.

2. Preferences cannot be defined over goods that include mental states in their description. The theory cannot employ, for example, good A defined as of 5 minutes of ecstasy or good C defined as 10 minutes of boredom. Using these sorts of goods does not allow RCT to bypass the EPOM, for we would now need to solve the EPOM to know which goods individuals have or would choose.

3. We must distinguish between the *accessibility* of mental states and their *necessity* for ICWs. If RCT is to bypass the EPOM, it must argue *not* that mental states are inaccessible, but that information about them is unnecessary to ICWs. For if RCT allows that mental states are necessary for ICWs, but eschews them only because they are inaccessible, then it provides no principled objection to the idea that mental states are a component of well-being, nor any principled objection to the value of addressing and solving the EPOM. This implies that in order to address the specific issue of necessity and disentangle it from problems of accessibility, *we are justified, when assessing RCT, in assuming away problems of accessibility*. We will therefore for the rest of this section assume that everyone's mind is transparent, that is, that we have the ability to read each

others' minds. The challenge for RCT, if it is to bypass the EPOM, is to show that even if we had the ability to read each others' minds--i.e. know whether people are happy, sad, bored, ecstatic, satisfied, in despair, etc.--the information revealed by this ability is unnecessary, because we wouldn't need it if we already have information about how people would choose.

4. If the transparency of minds or the ability to read minds is to have no value for ICWs given information about how people would choose, *it is necessary for RCT to hold that mental states are intrinsically irrelevant to well-being* (it is of course conceptually possible for mental states to be contingently relevant to well-being to the extent that they carry information about how people would choose, but we shall ignore this possibility). Consider John and me at our respective best choices A and B, who the zero-one rule considers equally well-off. Let us assume, furthermore that John is depressed at A while I am quite happy at B (we have this information because we are mind readers). Is such information about John's depression and my happiness intrinsically irrelevant? In particular, does such information about John's depression and my happiness give us reason to believe that the zero-one rule overestimates John's well-being relative to mine? The zero-one rule must answer in the negative. It must declare that facts about happiness, depression, life satisfaction, etc. do not matter at all and can never matter for well-being. For to allow them to matter would imply that I may be better off than John, even when we are both at our best choices, and this is a rejection of the zero-one rule.

Now someone who hopes that RCT might bypass the EPOM might object to this previous example, claiming that it is self-contradictory. He or she might claim that John's best choice, either by definition of best choices or by implication of the fact that it

is a best choice, cannot leave him depressed. But someone who wishes RCT to bypass the EPOM has no access to this claim. For depression is a mental state, involving a phenomenology of deep and persistent sadness. To make any claims about the relationship between choices and mental states is to imply both that there is something lawlike about mental states and that these laws are knowable, but this in turn assumes that we can successfully address the EPOM, which is precisely what RCT was supposed to render unnecessary. To successfully bypass the EPOM, RCT cannot claim that it is impossible for John to be depressed at his best choice, only that were it possible for him to be depressed at his best choice, his depression would be irrelevant.

5. I have just concluded that in order for RCT to bypass the EPOM, it must hold that mental states are intrinsically irrelevant to well-being. But a consequence of this conclusion is that if choices have an introspectively discernible impact on future mental states, then this impact *cannot be* part of the reason that choices are relevant to well-being. For example, if John would choose coffee over tea because the taste of coffee will give him more pleasure than the taste of tea, then the fact that he gets more pleasure from coffee rather than tea as a result of his choice *cannot be* the reason he is better off with coffee than with tea. Or for another example, if John would choose to take an anti-depressant over not taking it, and if the anti-depressant would cure him of a deep sadness, then the fact that the anti-depressant would cure of him of the deep sadness cannot be the reason he is better off with the anti-depressant than without. Or for yet another example, say that John would choose to undertake quixotic life project over not doing so, and say also that undertaking a quixotic life project gives him a giddy feeling that he values over many different kinds of sensation, then the giddy feeling that he gets from undertaking a

quixotic life project is not part of the reason he is better off when he is undertaking a quixotic life project than when he is not. For if that were the reason, then the relevance of choices to well-being would depend on the relevance of mental states to well-being. But this is ruled out by our previous conclusion. Therefore the reason choices are relevant to well-being can never be because of the mental states that would result from these choices.

4. The evaluative attitudes, value, and what makes a life go best

I have just discussed the claims of the zero-one rule, and the interpretations of it that we should adhere to if it is to provide us a way of bypassing the EPOM. Now I shall enumerate reasons to reject the zero-one rule, and therefore reasons to reject that the view that RCT allows us to bypass the EPOM.

1. The first and simple reason involves the very definition of choices. It seems that a definition of choice must include at least two elements, the first is some externally observable behavior (e.g. uttering "I'd like coffee"), and the second some psychological processes or phenomena that we assume or infer accompany that behavior. These psychological processes or phenomena might include a phenomenology of conscious will, or authorship, or executive agency, or deliberation. It is by specifying these psychological processes, and inferring the occurrence of choices only when the observed behavior is presumed to be accompanied by these psychological processes, that we can definitionally and empirically distinguish choices from other externally observable behaviors such as reflexes, automatisms, and sleepwalking, all of which we in turn conceptually and empirically distinguish from choices by the fact that these are *not* accompanied by those same psychological processes, or by the fact that they are

accompanied by different ones. Inferring the occurrence of choices therefore requires inferring the occurrence of the psychological processes distinctive of choices, and therefore requires making inferences about the contents of other minds.

To claim that a theory based on choices allows us to bypass the EPOM is therefore to play a kind of shell game. "If we have information about choices", this argument might go, "we would not need information about mental states." The problem with this argument is that information about choices necessarily contains information about the mental states that are definitional of or distinctive to choices. To say that John chose coffee is to say not only that John manifested some behavior (uttered "I'd like coffee, please"), but to say as well that a particular kind mental process accompanied this behavior (he resolved inside his head to have coffee), or to infer or assume that it did.

2. The second more important reason is philosophical. It says that any plausible theory of value, or of well-being, or of the reason that choices are relevant to well-being, must link both choices and well-being to the *evaluative attitudes*, those mental states that associate outcomes or behaviors with positive or negative value. Intimations of this philosophical reason are already familiar to the student of basic choice theory when he or she is warned about what can and cannot be concluded from information about preferences. For example, remember our assumptions that John's best and worst choices are A and C respectively, and that my best and worst choices are B and D respectively. The student of basic choice theory must be studiously warned by the instructor that no conclusion can be drawn about how much happier, pleased, or gratified either John or I would be to get our best instead of our worst choices. Nor therefore can any conclusion be drawn about the relative increase in happiness, pleasure, or gratification that John and

I would get from having our best rather than worst choices. Or assume that both John and I are indifferent between the two goods E and F. The student must be warned against drawing the conclusion that John derives as much happiness, or pleasure, or gratification from E that I get from F. Nor, remember, can we trick our way towards such conclusions by defining the goods in terms of the mental states, for example, by defining John's best choice A as the state of being profoundly happy and worst choice C as being in a state of deep depression, for this would simply reintroduce the EPOM in another form, that of not knowing what goods people possess or choose. The fact that the student needs to be warded away from those conclusions implies that the student may have an untutored intuition that well-being is intimately related to having certain kinds of positive or negative feelings towards our lives or its various features--feelings that include sensual pleasure or pain to be sure, but more generally also include happiness or sadness, a sense of purpose or aimlessness, a sense of connectedness with others or a sense of isolation, a sense of meaningfulness or meaninglessness, of eagerness or ennui, etc--and that the instructor must remind the student that choices are only imperfectly related to such feelings. This unbreakable link between well-being and positive and negative feelings towards various aspects of our lives is ultimately the fundamental objection to the attempt to make ICWs through RCT in a way that bypasses the EPOM.

A traditional way to categorize mental processes is into the three categories of cognition, emotion, and motivation[4], each of which corresponds to a specific kind of dichotomy. Cognition attempts to answer true or false questions, emotions involve valuations of good or bad, and motivation involves tendencies to either approach or retreat. The emotions are sometimes also called the evaluative attitudes, and its

valuations of good or bad can involve varying degrees of judgment or deliberation. The latter two elements of emotion and motivation are sometimes jointly called the affective states. The traditional dichotomy of reasons versus passion involves the dichotomy between cognition and affect. It is also part of the traditional understanding of both emotions and motivation that each distinct kind of emotion or motivation admits of a scalar measure of intensity or quantity, sometimes called *hedonic tone*. It is this hedonic tone that allows us to make claims such as that one person is happier, or more anxious, or more excited, or more tempted than another.

How are the evaluative attitudes related to well-being? I shall discuss this question in terms of Parfit's[5] three most plausible accounts of what makes a life go best: Preference Hedonism (PH), Success Theory (ST), and the Objective List Theory (OLT). In his discussion of these three theories, he employs the language of desire satisfaction, but I will use the language of the evaluative attitudes, which I believe involves no distortion. The first two of these, PH and ST, involve a normative *anti-realist* theory of value while the last, OLT, involves a *realist* theory of value. *Anti-realism* about value says that all value in the universe ultimately flows from the evaluative attitudes, and that an individual's well-being must ultimately be constituted by his or her evaluative attitudes towards his or her life. PH and ST differ in that ST requires that my evaluative attitudes not be based on any beliefs that suffer from errors of non-normative fact, while PH allows such evaluative attitudes. The two theories come apart when I have beliefs suffering from errors of non-normative fact that prevent me from introspectively discriminating between two different lives toward which, if I knew the truth, I would hold different evaluative attitudes. A classic example is the case where I have a preference

against being deceived. Consider life B where I am not being deceived, and life D where I am being deceived but falsely believe that I am not. Since my false beliefs in B prevent me from introspectively discriminating between B and D, I will actually hold identical evaluative attitudes towards both of them. According to PH, I am therefore equally well-off in B and D. But according to ST, if my beliefs suffered from no errors of non-normative fact, then I would recognize that I have succeeded in my aims in B but failed in D. If my beliefs suffered from no errors of non-normative fact, I would have worse evaluative attitudes towards D than I do towards B, implying that I am less well off in D than in B.

Despite this difference between PH and ST, both hold that it is ultimately the evaluative attitudes that matter for well-being. Indeed they are in complete agreement about judgments of well-being under the circumstance when beliefs suffer from no errors of non-normative fact. But even the fact that our beliefs can suffer from errors of non-normative fact does not diminish the relevance of the evaluative attitudes. For the badness of having false beliefs must depend on our actual evaluative attitudes towards the states of affairs that having those false beliefs prevents us from discriminating from other states of affairs. Imagine that despite my best efforts to recruit true friends, these efforts fail and I end up having some false friends. According to ST, I am worse off than I think I am. But now ask: *how much* worse off? Presumably for the anti-realist ST, how much worse off I am must ultimately be determined by my evaluative attitude towards the state of having false friends, which is an actual evaluative attitude that I possess toward that state even when I am unable to empirically distinguish it from other states. In other words, the fact that I cannot distinguish whether I live in a particular state or not does not

imply that I don't have evaluative attitudes toward that state, nor does it imply that that evaluative attitude doesn't determine how worse off I am for being in such a state. My evaluative attitudes determine my true well-being even if I do not know my true well-being. An impartial observer who believes in ST, who must assess my well-being, and who does not labor under the same erroneous beliefs that I do, still needs to know my evaluative attitudes towards the states that my false beliefs prevent me from distinguishing from other states.

The third theory, OLT is *realist* about value. That is, it claims that there are normative facts about what evaluative attitudes we *ought* to have about our lives, and such normative facts are irreducible to facts about our evaluative attitudes. Examples of such normative facts might be that I ought to care about my distant future as much as I care about my near future, or that I ought not take pleasure from others' needless suffering. According to this theory, I ought to make my evaluative attitudes conform to these normative facts, and this obligation does not stem from some prior evaluative attitudes, but from the brute reality of the normative facts themselves. A standard example that differentiates, say ST from OLT is a hypothetical Caligula who takes pleasure from others' suffering but who does not suffer from beliefs that involve errors of non-normative fact. For ST, this hypothetical Caligula experiences pleasure in the absence of false non-normative beliefs, and may therefore be thought of as well-off. But for OLT, this hypothetical Caligula's pleasure is normatively unreasonable. And to have normatively unreasonable evaluative attitudes is to be worse off, therefore this hypothetical Caligula is worse off than ST implies.

According to OLT, and in contrast to PH and ST, it is not our evaluative attitudes that

determine our well-being, but their normative reasonableness. But it is also obvious that this makes information about the evaluative attitudes *no less necessary* for ICWs. Information about the evaluative attitudes is necessary *when* they are reasonable. This information is also necessary in order to be able to assess *if* they

make identical choices may have different evaluative attitudes, and people with identical evaluative attitudes may make different choices. Consider two people, Mack and Jenny, who have identical preferences in the sense that they would make the same choice given any pair of alternatives. That is, they would both choose M in a choice between M and N, and they would both choose O in a choice between O and P, and so on, for all possible pairs of choices. Do we have any reason to believe that they would have identical evaluative attitudes at *any* of the objects of their choices? That is, do we expect that they would be equally happy at O? Or equally happy at P? Or indeed at any other good? Or assuming they had best and worst choices, that they at least have identical evaluative attitudes at their best choices and identical evaluative attitudes at their worst choices? The answer to this must be no. Indeed this is precisely the type of conclusion that the student of basic choice theory is being warned against. If Mack were just persistently happier, and Jenny persistently sadder, and if Mack's various choices are associated with more or less happiness but if Jenny's various choices are associated with more or less sadness, this would be enough to frustrate the sufficiency of choice information. Or if Mack's happiness is much more sensitive to the fulfillment of his preferences than Jenny's, this would also be enough as well. Thus it seems clearly false that information about choices necessarily contains all the necessary information about the evaluative attitudes. I also conclude from this that it is also clearly false that two people who are at their respective best (worst) choices must be equally well off, and better off (worse off) than any other people who are not at their respective best (worst) choices.

Choices and evaluative attitudes can come apart for numerous reasons. One important reason is that some evaluative attitudes are *not directed towards objects of choices*. An

example consists of moods, which may be powerful and overwhelmingly positive or negative, but are not directed towards an object of choice and therefore is relatively insensitive to choices. Or one may be terribly depressed at the prospect that the passage of time is an illusion, that the human species will go extinct, or that the universe will end in a heat death. Another important reason is the so-called thesis of *separation between evaluation and motivation*[6] which says that what we value may fail to motivate us to act, and what motivates us to act may not be valued by us. Some classic exemplars of the separation thesis are weakness of the will, apathy, and perversity, the first involving being motivated by what we do not value, the second involving the inability to be motivated by what is valued, and the last involving being motivated by something *because* we do not value it. Behavioral economists and psychologists emphasize the impact on people's choices of biases and heuristics and of their poor predictions of how their future mental states will be affected by these choices. In more mundane and everyday settings, the relationship between people's choices and their evaluative attitudes towards the objects of their choices can be confounded by *other* alternately competing or complementary values or motivations such as a desire to deceive, or to conform, or to please. In more extreme but no less conceptually relevant scenarios such as automatisms, sleep walking, or robotic simulation, what has the public appearance of a choice may be untied to any evaluative attitudes at all. For all these reasons, I shall henceforth take as unassailable fact the contingent relationship between choices and the evaluative attitudes. To summarize, information about the evaluative attitudes are necessary for ICWs, and since choices and these attitudes can come apart, information about choices is not sufficient for ICWs. Parenthetically, the same line of argument can be used to conclude

that there is also only a contingent relationship between a person's evaluative attitudes and his or her other publicly observable attributes such as income, education, health, legal rights, and so on. Therefore information about such attributes is also insufficient for ICWs.

The second thing to be said about the claim that information about choices might contain all the necessary information about the evaluative attitudes is that this claim is ruled out for those who would hope that RCT can find a way to bypass the EPOM. For to try to establish this claim is to imply that we have sufficient objective knowledge about the evaluative attitudes and their relationship to choices to establish the truth of the claim. And again, this is precisely the kind of determination that RCT must show to be unnecessary if it is to bypass the EPOM. To bypass EPOM, RCT really must claim that the evaluative attitudes are irrelevant to well-being, and irrelevant to the relevance of choices to well-being. But then this brings us to a question that RCT must answer: if it claims that neither choice nor well-being is related to the evaluative attitudes, and if it claims that choice is nevertheless still relevant to well-being, *what could the reason for the relevance of choice to well-being be*? Why should we think that people become better off when they get what they would choose? The reason for the relevance of choice to well-being is easily given if both are linked to the evaluative attitudes--the relevance of choices to well-being would be neither intrinsic nor analytic, but rather contingent upon the relationship between choices and the evaluative attitudes--but the reason for relevance is wholly obscure if neither well-being nor choices is linked to these attitude. It just seems to me that there is no non-tautologous reason to hold choice to be relevant to well-being when both are divorced from the evaluative attitudes. A defender of RCT may

respond that this is true, but that it is no criticism of RCT, for RCT *tautologously defines* well-being in terms of choices, i.e. we should understand the zero-one rule as *defining* well-being from choices in such a way that makes ICWs possible. But this assertion does not provide RCT with an escape from our criticism. For though it involves no logical error to tautologously define well-being in terms of choices, I can still justifiably demand *a reason in favor of asserting the tautology*, and this reason cannot be the tautology itself. And I simply cannot see what that reason could be. The idea that choices are relevant to well-being independently of the evaluative attitudes is potentially incoherent, wholly obscure in rationale, and finds no defense in any of our best theories of value or of what makes a life go best. One reason that might be given in favor of invoking the tautology, indeed a reason with historical explanatory power, is that we assert the tautology because the EPOM is intractable and so we have no choice but to invoke it. But if this is the reason for asserting the tautology, then it is not a reason that allows RCT to bypass the EPOM, for this is a reason that only shows mental states to be inaccessible, not unnecessary for ICWs. If not for the relationship of choices and well-being to the evaluative attitudes, there would be no reasons to believe or assert that choices are relevant to well-being at all.

A potential critic may respond to our point about how the evaluative attitudes and choices can come apart by saying that RCT holds only for "fully rational" people, i.e. people for whom there is a well-ordered relationship between the evaluative attitudes and choices so that they do not come apart. The response to this potential criticism, in the spirit of the previous paragraph, is that to posit the possibility or coherence of a well-ordered relationship between the evaluative attitudes and choices is to assume not only

that the evaluative attitudes have a systematic and lawlike nature that can link up with choices, but that we can describe this link and empirically ascertain whether or to what extent this link holds. But this assumes that we can confront the EPOM, which is again what RCT is supposed to help us bypass.

6. Against extended preferences

The other major attempt to perform ICWs within RCT is called the *extended preferences* approach, most closely associated with Harsanyi[7] and which goes as follows. Consider the two outcomes "me living with A" and "me living with B". RCT assumes that I have preferences of the usual sort over such outcomes. These usual preferences consist of judgments of the form "I prefer to be me living with A over me living with B". But now consider the two outcomes "me living with A" and "John living with B". The extended preferences approach assumes that I have preferences over such outcomes as well, and such extended preferences consist of judgments of the form "I prefer to be me living with A than to be John living with B" or perhaps "I prefer to be John living with B than to be me living with A". These preferences are called extended because they are defined over extended outcomes, that is, outcomes that specify not just goods or life circumstances but also the identity of the person who is to possess those goods or inhabit those life circumstances. I think it is fairly obvious that in our everyday lives, we do have extended preferences. For example, we seem to be perfectly capable of saying things like "I wouldn't swap places with him for a million dollars" or "I would swap places with him in an instant." Such statements seem to imply that I have extended preferences over combinations of personal identities and life circumstances. But the extended preferences approach not only assumes that we have extended preferences. It

also assumes, much more strikingly that each of us is capable of having *perfectly sympathetic* extended preferences. Perfect sympathy is the assumption that my extended preferences, when limited only to outcomes that involve only John, for example the outcomes "John with A" and "John with B", coincide with John's own preferences for those outcomes.

Given the formal assumption that each of us has perfectly sympathetic extended preferences, the approach attempts to derive the surprising formal conclusion that we would all, despite our different non-extended preferences, nevertheless have identical extended preferences. If such a surprising conclusion indeed followed, then a route to ICWs would exist. For if this unique set of extended preferences were to satisfy the expected utility assumptions and Bernouilli's hypothesis, then extended preferences over gambles involving gains and losses to different people can be used to derive the values of ratios of changes in their well-being. However, Broome has argued that this surprising conclusion does not actually follow, and that Harsanyi's belief that this surprising conclusion could be reached was due to a technical mistake. This argument between Broome and Harsanyi is not relevant to my topic, and so I will leave the reader to consult the relevant literature for more detail[8]. My criticism of the extended preference approach involves the meaning of its assumption that each of us has extended preferences that exhibit perfect sympathy. For there seems to be only two possible meanings to perfect sympathy, and one does too little while the other does too much. First, if all perfect sympathy means is that my extended choices coincide with John's non-extended choices, then our entire argument against the zero-one rule applies to it. Knowing how John would choose is not sufficient for knowing about John's evaluative attitudes and therefore

his well-being. But second, if what perfect sympathy means is that I can put myself in his shoes, inhabiting his subjective point of view, experiencing his mental states and evaluative attitudes, then assuming perfect sympathy is simply *begging the question*. The fundamental challenge has always been how to solve the EPOM. One cannot solve this by just assuming that we can solve it, which is what the assumption that our extended preferences display perfect sympathy amounts to.

Broome himself[9] offers a non-preference-based theory of ICWs, which is worth discussing briefly. Assume that we have a theory of well-being according to which the goodness of a life, any life, is wholly determined by some set of features of a life F so that any two lives that are identical with respect to these features must by definition be equally good. Assume further, following our discussion of the zero-one rule, that ratios of changes in well-being are well-defined *intrapersonally*, so that there is intrapersonal comparability of levels and differences in well-being. Finally call the set of lives that John might live X and the set of lives that I might live Y. Then if some life *x1* of John's from the set *X* is identical with respect to features F as some life *y1* of mine from the set *Y,* then John's living life *x1* and my living life *y1* are equally good. And if John and I have some other lives *x2* and *y2* respectively, which are also identical with respect to the features F that determine well-being, then John's living life *x2* and my living life *y2* are equally good. Since we now have two points of utility at which John's and Paul's well-being are equal to each other, intrapersonal ratio-scale comparability accomplishes the rest. I do not argue with this theory, for it seems correct. I only claim here that our previous discussion implies that the features F that are relevant to well-being include the evaluative attitudes. Therefore to know whether two lives are identical with respect to

features F requires comparing their evaluative attitudes. That is, it requires addressing the EPOM.

I conclude, contra RCT, that ICWs necessarily require interpersonal comparisons of individuals' evaluative attitudes towards their lives and therefore a solution to the EPOM.

7. A proposed solution to the EPOM and redefinition of ICW

The EPOM involves whether and how I might have objective knowledge of your subjective mental states. Some might believe that the chasm between me and your mental states just cannot be bridged, that the realm of your mind is so essentially private that there are no empirical facts accessible to me that can shed light upon it. This skepticism might in turn be inspired by an either implicitly or explicitly dualist view that the mental and physical realms are distinct from and independent of each other, and that the operations of the mind are distinct from and independent of the operations of the brain, a view most famously held by Descartes[10]. But a dualist view involving independence of operations of the mind from operations of the brain is now believed to be false and untenable by the vast majority of neuroscientists and philosophers of mind. In contrast to the *contingent* relationship between mental states and behavior, it is now widely believed that mental states and brain states are *non-contingently* related.

All empirical evidence and respectable scientific and philosophical approaches to the study of mind now support what we might call a *type correspondence between mind and brain*. This is the view that every type of mental state or event corresponds to a type of physical brain-based state or event such that a specific mental event of a particular type occurs at some particular time and place if and only if a specific brain event belonging to the corresponding physical brain-based type also occurs at that time and place. So for

example, consider the type of mental event that is the feeling of pain. To the best of our current scientific knowledge, the type of brain event to which it corresponds involves the activation of the Anterior Cingulate Cortex[11]. John will experience pain at 8 am on October 15, 2010 in Cambridge, MA if and only if his ACC activates at exactly those same coordinates. The type correspondence view says that analogous results hold for all types of mental events.

The type correspondence view says that any introspectively discernible feature or structure of a mental state must correspond to some objective feature or structure in the corresponding brain state. For example, if pain varies in intensity, there must be some variation in the activation of the ACC that corresponds to it, so that a pain of a certain intensity occurs if and only if an activation of the ACC of a certain degree occurs. Indeed *any* differentiation between two instances of the same type of mental state (e.g. two pains of slightly different intensity) no matter how minute must correspond in a systematic way to *some* differentiation between their corresponding brain states.

This implies that two different mental states cannot correspond to the same brain state, for this would imply the existence of variation in the mental that is uncorrelated with variation in the physical. Two different mental states *must* correspond to at least two distinct brain states. This implication might be called *minimal supervenience of mind on brain*[12] (Davidson 1970 introduced the concept of supervenience into the literature on the mind-brain relationship). A way of restating minimal supervenience is to say that every specific brain state can correspond to no more than one mental state. Yet another way of stating it is that two people with different mental states must have different brain states, and two people with identical brain states must have identical mental states.

Type correspondence is a claim about how our mental lives are--contra Descartes--inseparable from the workings of the brain and that there is no such thing as a free-floating mental state hovering above and unanchored to the physical world. Try to list every possible mental event: from seeing an apple, to wanting an ice cream, to feeling pain or pleasure, to choosing your entrée from a menu, to planning for retirement, to hating someone with a passion, to plotting a murder, to assessing the relative weight of competing practical reasons, to committing or identifying one's self with a particular set of practical reasons. All such events supervene on some physical event in the brain, an event that can be described exclusively in terms of neurons firing, neurotransmitters bonding, electricity, chemistry, and so on. These mental events cannot happen unless the subvening physical events also happen. And not only is the mental linked to the physical, but the link between them is lawlike, systematic, exceptionless, and discoverable through scientific investigation.

The belief in type correspondence and minimal supervenience is arrived at not a priori, but on the basis of overwhelming and pervasive empirical evidence of correlations between mental states and brain states. In our ordinary lives these correlations reveal themselves in the impact of brain trauma, injury, and disease on our mental states, and in the impact of changes in brain chemistry brought about by alcohol, medicines, and drugs on our mental states. In scientific settings, these correlations are revealed in experimental manipulations and stimulation of the brain states of study subjects and their self-reports of their resultant mental states. Examples of empirical correlations between mental and brain states include the experience of pain and the aforementioned activation of the brain's Anterior Cingulate Cortex (ACC)[13]; the experience of pleasure and flows of

neurotransmitters such as dopamine and serotonin[14]; the conscious exercise of self-control and the activity of the anterior prefrontal cortex[15]; religious feeling and the stimulation of medial temporal lobe[16]; social emotions such as empathy, guilt, shame, and compassion and activity in the ventromedidal prefrontal cortex[17]; the recognition of faces and fuciform gyrus area[18]; the experience of intending to act and activity in the frontal and parietal brain areas[19]; lying and the bilateral ventrolateral pre-frontal and anterior cingulate cortices[20]; and fear and the amygdala[21] to give a few examples.

Belief in type correspondence and minimal supervenience is a relatively undemanding requirement. It does not require taking a stand on controversial issues such as whether mental states are actually just *identical* to brain states, or whether we should think of mental states as *emergent properties* of brain states, or whether the correspondence involves *causation* from brain states to mental states, or whether the same type of mental state is *multiply realizable* in different types of physical states in different species or lifeforms, or even whether some modified Cartesian dualist view in which the mental realm is *metaphysically distinct from but mirrors* the physical realm is correct. It certainly doesn't require taking a stand on the correct solution to the great mystery of consciousness, i.e. the mystery of how third-person phenomena like brain states can give rise to first-person phenomena like mental states. But though this belief is undemanding, it is also in a sense unforgiving in that *any* reasonable stance on any of these controversies *must* accept type correspondence and minimal supervenience, or at least accept that they hold in our species and in our actual universe. Any view that denies type correspondence and minimal supervenience is occult. Type correspondence and minimal supervenience are, to be sure, not compatible with all metaphysical views about the

mind-brain relationship: it is, for example, incompatible with Chalmers' Zombie Hypothesis[22], i.e. that it is logically conceivable and therefore actually possible for a brain to exist that is in all physical respects exactly like yours but which has none of your mental states.

Earlier in this paper, I said that the relationship between mental states and behavior, and a fortiori between the evaluative attitudes and choices was *contingent*. By this, I mean that there is no if-and-only-if relationship between occurrences of the former and occurrences of the latter. For example, there are some occurrences of pain that are not accompanied by an utterance of "Ouch!" and there are some utterances of "Ouch!" that are not accompanied by occurrences of pain. In contrast, type correspondence says that an if-and-only-if relationship *does* exist between occurrences of mental states and brain states. This is what it means for the relationship between them to be *non-contingent.* Now it may be asked, what is the epistemic basis for the claim that there is an if-and-only-if relationship between mental and brain states? Don't scientists and clinicians who try to uncover the form of this relationship in laboratories and experimental settings also ultimately have to rely on people's behaviors and testimony, and doesn't what I've just said imply that *those* behaviors and testimonies are *not* in an if-and-only-if relationship to mental states? The answer to this is yes. The epistemic basis for the claim of type correspondence has to be that there are certain privileged circumstances--such as in a laboratory or in experimental settings, or when a researcher is using a well-designed survey questionnaire and research protocol designed to minimize respondent biases, in contrast to non-privileged circumstances such as the messy everyday world of social interaction--in which behaviors and utterances are more trustworthy indicators of mental

states, and that under such privileged circumstances, the data support the inference of type correspondence.

Type correspondence implies that the operations of the mind are linked in lawful and exceptionless ways to the operations of the brain. This provides a principled solution to the EPOM because brain states are objectively observable in a way that private and subjective mental states are not. Variations in brain states shed objective light on variations in subjective mental states, including their intensities. So information about individuals brain states, plus scientific information about the mental states to which they correspond, allow us to make inferences about mental states, particularly the evaluative attitudes and their intensities. My argument is that information about brain states, plus information about the mental states to which they map, give us a principled solution to the EPOM. However, in practice, the quality of our information about both brain states and the mental states to which they map are determined by the extent of scientific progress. So the practical utility of the solution proposed here is contingent on the extent of scientific progress. There are already a few narrowly defined practical situations where inferring mental states through brain states is an option. Farah[23] discusses the cases of patients in minimally conscious states and of non-human animals. Spence[24] discusses deception and truth-telling in the law. Neuroscience may be able to find even more direct solutions to the EPOM than we have sketched here. As Ramachandran and Blaskeslee[25] point out, no scientific law rules out the possibility that we may one day be able to record your brain states and reconstruct or transfer them to other brains or machines. If the technology some day exists to plug your brain into mine, or to reconstruct some of your brain states in mine, then I could directly experience at least

some portion of your mental states (this may never allow me to experience the totality of your mental states, since presumably I would need to preserve some of my brain states against being overwritten by yours so that it remains *me* who is experiencing your mental states). This would show the inaccessibility of other minds to result not from some intrinsically unbridgeable chasm but a technology gap. Such a technological possibility would allow me a partial knowledge of your mind, not through an inference of your mental states from your brain states, but through direct experience of your mental states. But this would be getting ahead of ourselves. On the whole, we are in the early days of neuroscience, and it may well take centuries or even longer before science gives us the ability to read minds with as much precision as might be necessary for normative decision-making that ICWs are supposed to inform. In the meantime, we can do no better than continue as before making inferences about mental states from contingently related behavior and choices, rather than from non-contingently-related brain states.

The evaluative attitudes are a particular kind of mental state, indeed a proper subset of them. Type correspondence therefore implies that there are brain states that correspond to the evaluative attitudes, and that physical aspects of those brain states correspond to every introspectively discernible aspect of these attitudes, in particular their intensity. Neuroscience can be expected to improve our ability to map the intensities of the evaluative attitudes onto specific aspects of the brain states that subvene them. An example of this is described Smith[26] who reports on research that seems to have discovered a neural signal, the duration of a particular kind of low-frequency brain wave, that correlates with the intensity of pain, so that "the more pain that is experienced, the longer the waves last". If type correspondence is true, analogous results must be true for

every kind of evaluative attitude. Hence for any given evaluative attitude, type correspondence tells us that we can in principle tell whether any two individuals are experiencing equal intensities of that attitude, and if not, who is experiencing it more strongly.

One can have evaluative attitudes towards many things. I can have evaluative attitudes towards any aspect of my life, small and large: I can really dislike a pain in my shoulder, but really enjoy my family life. Indeed and more generally, I can have evaluative attitudes towards any aspect of the universe: I can feel sad at the thought that the universe and all life in it might end in a heat death. I can have evaluative attitudes towards the fact that my life or more generally the universe is one way rather than another: I can feel regret that I am not a better singer, or be relieved that life on Earth came into being. I can have evaluative attitudes towards my evaluative attitudes: I might feel embarrassed that I like James Bond movies. I may have evaluative attitudes that don't seem directed at anything at all: I can just be happy, or just be worried. But from the point of view interpersonal comparisons of *well-being* specifically, it seems that it must be my evaluative attitudes towards particular aspects of my life that matter. In the literature on the psychology of well-being (see Urry et. al. 2004 for a review), conceptualizations of well-being seem to fall into two camps. The first, exemplified by Diener's work[27], is the hedonic approach in which well-being has four elements (i) satisfaction towards one's life as a whole, (ii) satisfaction towards important life domains such as work, (iii) frequent pleasant emotions, and (iii) infrequent unpleasant emotions. The second, exemplified by Ryff's work[28], is the eudaimonic approach in which well-being depends on the degree of one's positive evaluations of (i) one's own self and life history, (ii) one's relationships

with others, (iii) one's degree of autonomy, (iv) one's mastery of one's environment, (v) one's life purpose, (vi) one's scope for personal growth. Thus whether hedonic or eudaimonic, well-being is tied to various emotions and evaluative attitudes towards various domains of life.

And it is perhaps worth noting explicitly and separately that I can have evaluative attitudes towards changes or prospects of changes in my life. It is these attitudes towards changes and prospects of changes that allow for interpersonal comparisons of changes in well-being. To see why this is plausible, consider someone who believes that intrapersonal comparisons of changes in well-being are somehow related to that person's choices over gambles. If choices over gambles are in turn reflective of that person's evaluative attitudes towards the prospects that are contained in those gambles--as they must be if such choices are to be more than reflexes or automatisms or mindless acts--then intrapersonal comparisons of changes in well-being must be related to people's evaluative attitudes towards prospects of changes. Interpersonal comparisons of changes in well-being must in turn be related to comparisons of different people's evaluative attitudes towards prospects of such changes.

8. The evaluative attitudes and well-being: the many versus the one

I have so far argued for what might be called the interpersonal comparability of the evaluative attitudes. I have also claimed that, subject to the conditions of accuracy of beliefs and reasonableness of attitudes, well-being supervenes on the evaluative attitudes. But these claims stop short of providing a more complete picture of the relationship between the evaluative attitudes, which seem to be multifarious, and well-being which has the appearance of being a unitary concept. But addressing this distinct

and deeply difficult question is beyond the aims of this paper. What would be most convenient is if there were a kind of summary or aggregative or master evaluative attitude whose intensity can serve as an index of an individual's over-all well-being. If there were such a master attitude, a measure of its intensity could serve as the target of all efforts aimed at promoting the personal good. Recent research into the empirical measurement of self-reported hedonic and eudaimonic accounts of well-being seem to be aimed at identifying and measuring such a summary attitude. Indeed Urry et. al.[29] report on research that is embarked upon measuring the brain states that subvene such summary attitudes. But it is a far from settled issue in either science or philosophy whether the Self has sufficient unity to guarantee that the multifarious evaluative attitudes fit into a coherent whole or allow for unproblematic summary. There is in fact a growing literature that points in the other direction, claiming that the Self is fragmented rather than unitary[30]. If well-being is an attribute of selves, then the fragmentation of selves may imply the fragmentation of well-being. This in turn may mean there is nothing to interpersonal comparisons of well-being *other than* interpersonal comparisons of this or that evaluative attitude.

Notes

[1] See W.S. Jevons, The Theory of Political Economy (London: MacMillan, 1871); L. Robbins, An Essay on the Nature and Significance of Economic Science (London: Macmillan, 1932); R. Hare, Moral Thinking (Oxford: Oxford University Press, 1981); J. Harsanyi, Interpersonal Utility Comparisons (The New Palgrave Dictionary of Economics Online. Palgrave Macmillan. 01 June 2009 <http://www.dictionaryofeconomics.com/article?id=pde2008_I000196> doi:10.1057/9780230226203.0841; M. Farah, Neuroethics and the problem of other minds: implications of neuroscience for the moral status of brain-damaged patients and non-human animals. Neuroethics 1 (2008): 9-18.

[2] discussed and criticized at length in D. Hausman, The impossibility of interpersonal utility comparisons. Mind 104(1995): 473-90.

[3] J. Broome, Weighing lives (New York: Oxford University Press, 2004)

[4] B. Parkinson and A. Colman, Emotion and motivation (London and New York: Longman, 1995)

[5] D. Parfit, Reasons and persons (Oxford: Oxford University Press, 1984)

[6] see S. Tenenbaum, Appearances of the good: an essay on the nature of practical reason (Cambridge: Cambridge University Press, 2007)

[7] J. Harsanyi, Rational Behaviour and Bargaining Equilibrium in Games and Social Situations (Cambridge: Cambridge University Press, 1977)

[8] see J. Broome, A cause of preference is not an object of preference. Social choice and welfare, vol. 10 (1993): 57-68 and J. Broome, Can there be a preference-based utilitarianism? in Fleurbaey, Salles, and Weymarck, eds. Justice, Political Liberalism,

and Utilitarianism. (Cambridge: Cambridge University Press. 2008)

[9] J. Broome, Weighing Lives, (New York: Oxford University Press, 2004).

[10] R. Descartes, Discourse on Method and the Meditations, trans. F. E. Sutcliffe (Penguin, 1968)

[11] Farah, Neuroethics and the problem of other minds.

[12] The concept of supervenience was introduced into the literture on the mind brain relationship in 1970 by D. Davidson, Mental events. Reprinted in Davidson, ed., Essays on Actions and Events. (Oxford: Clarendon Press, 2001)

[13] Farah, Neuroethics and the problem of other minds.

[14] R. Carter, Mapping the mind, (Berkeley and Los Angeles: University of California Press, 1998)

[15] N. Shamosh, C. DeYoung, A. Green, D. Reis, M. Johnson, A. Conway, R. Engle, T. Braver, and J. Gray, Individual differences in delay discounting: Relation to intelligence, working memory and anterior prefrontal cortex. Psychological Science 19(2008): 904-11.

[16] Carter, Mapping the mind.

[17] M. Koenigs, L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio, Damage to the prefrontal cortex increases utilitarian judgments. Nature 446 (2007) 908-11.

[18] G. McCarthy, A. Puce, J. Gore, and T. Allison, Face-specific processing in the human fusiform gyrus. Journal of Cognitive Neuroscience, 9 (1997): 604-609.

[19] P. Haggard, Conscious intention and motor cognition. Trends in cognitive sciences 9(2005): 290-5.

[20] S. Spence, S. The deceptive brain. Journal of the Royal Society of Medicine 97(2004):

6-9.

[21] J. LeDoux, Amygdala. Scholarpedia 3(2008): 2698.

[22] D. Chalmers, The Conscious Mind: In Search of a Fundamental Theory. (New York and Oxford: Oxford University Press, 1996)

[23] Farah, Neuroethics and the problem of other minds.

[24] Spence, The deceptive brain.

[25] V. Ramachandran and S. Blakeslee, Phantoms in the brain. (New York: Harper Perennial, 1998)

[26] K. Smith, Brain waves reveal intensity of pain. Nature 450(2007): p. 329.

[27] E. Diener, Subjective well-being: the science of happiness and a proposal for a national index. American Psychologist 55 (2000): 34-43.

[28] C. Ryff, Happiness is everything, or is it? Explorations on the meaning of psychological well-being. Journal of Personality and Social Psychology 57(1989): 1069-81.

[29] H. Urry, J. Nitschke, I. Dolski, D. Jackson, K. Dalton, C. Mueller, M. Rosenkranz, C. Ryff, B. Singer, and R. Davidson, Making a life worth living: neural correlates of well-being. Psychological Science. 15 (2004):367-72.

[30] P. Churchland, Self-Representation in Nervous Systems, Science 296(2002): 308-310.