
Munich Personal RePEc Archive

The Mirror-Neuron Paradox: How Far is Sympathy from Compassion, Indulgence, and Adulation?

Elias Khalil

Monash University

11. June 2007

Online at http://mpra.ub.uni-muenchen.de/3509/
MPRA Paper No. 3509, posted 12. June 2007

The Mirror-Neuron Paradox:
How Far is Sympathy from Compassion, Indulgence, and Adulation?

Elias L. Khalil[1]

ABSTRACT

The mirror-neuron system (MNS) becomes instigated when the spectator *empathizes* with the principal's intention. MNS also involves *imitation*, where empathy is irrelevant. While the former may attenuate the principal's emotion, the latter paradoxically reinforces it. This paper proposes a solution of the contradictory attenuation/reinforcement functions of fellow-feeling by distinguishing two axes: "rationality axis" concerns whether the action is efficient or suboptimal; "intentionality axis" concerns whether the intention is "wellbeing" or "evil." The solution shows how group solidarity differs from altruism and fairness; how revulsion differs from squeamishness; how malevolence differs from selfishness; and how racial hatred differs from racial segregation.

Keywords: Adam Smith; David Hume; Fellow-Feeling; Desire; Paris Hilton; Animal Rights; Comprehension; Understanding (empathy); Imitation; Status Inequality; Elitism; Authority; Pity: Obsequiousness; Racial Segregation; Racial Hatred; Rationality; Intentionality; Propriety; Impropriety; Revulsion; Stigler's Crankcase Oil Problem; Social Preferences; Altruism; *Assabiya* (group solidarity); Fairness; Theory of Evil (spite/malevolence)

JEL Code: D01; D64

[1] Elias.khalil@buseco.monash.edu.au Department of Economics, Monash University, Clayton, Victoria, Australia. The paper was supported by the Konrad Lorenz Institute for Evolution and Cognition Research (Altenberg, Austria). During my stay at the Konrad Lorenz Institute, I benefited greatly from the very generous comments and extensive conversations with Riccardo Draghi-Lorenz. I also benefited from the comments of Philippe Fontaine, Ulrich Krohs, Robert Sugden, and seminar participants at the Konrad Lorenz Institute, George Mason University (Center for Public Choice), and Monash University. The usual caveat applies.

The Many Faces of Fellow-Feeling

In her only published novel, *To Kill a Mockingbird*, Harper Lee tells stories about everyday life and racial segregation in a backwoods town in the Deep American South. The novel takes place in 1932, Maycomb County, Alabama. Tom Robinson is an African-American young man wrongly accused and, without one iota of evidence, convicted of raping Mayella Violet Ewell, a 19-year old white woman. In his way home from the fields, and responding to her requests, Tom helped Mayella over many months with chores in the yard without taking a penny from her. The prosecutor, Mr. Gilmer, leveled a barrage of questions as to why would Tom help the woman: "Why were you anxious to do that woman's chores"—with her father and seven children on the place? "You did all this chopping and work from sheer goodness, boy?"? "You're a mighty good fellow, it seems – did all this for not one penny?" [Lee, 1989, pp. 217].

Tom finally explained: "I felt right sorry for her, …." Sure enough, there are plenty of reasons to feel sorry for Mayella: her mother has long been dead, her father drank most of the relief check and abused her when drunk, and she was the oldest of so many younger siblings. But as soon as Tom uttered his words of fellow-feeling, he interrupted himself. He realized that he made a big mistake. Mr. Gilmer gleamed over his prize:

> `*You* felt sorry for *her*, you felt *sorry* for her?' Mr. Gilmer seemed ready to rise to the ceiling.
>
> The witness realized his mistake and shifted uncomfortably in the chair. But the damage was done. … nobody liked Tom Robinson's answer. Mr. Gilmer paused a long time to let it sink in [Lee, 1989, pp. 218].

Tom definitely damaged his case. How could he dare, a black person, feel sorry for a white

person?[2] If he felt any fellow-feeling, it should be the obsequiousness towards white people. And they, in return, would feel pity towards him. So, for Tom to claim that he felt sorrow for Mayella can only be interpreted by the white jury as pity, i.e., what they feel towards black people. Even if Tom's fellow-feeling is empathy, empathy entails status equality. It is obvious to anybody, given the institutional matrix of status inequality and elitism, Tom's motive cannot be empathy. It should not be surprising, therefore, that the jury found Tom guilty of rape.

This paper uses the term "fellow-feeling" as a primitive, i.e., as the most elementary unit of which more complex emotions are made. Fellow-feeling actually is so elementary that it has no meaning if abstracted from the institutional matrix, such as status inequality or racial segregation, as Tom's predicament illustrates. The institutional matrix allows fellow-feeling to take a multitude of recognizable faces. If one ignores the institution matrix, one would be perplexed as to why Tom's fellow-feeling worked against him. It would be naïve to assume that the exchange of fellow-feeling among agents has the same meaning irrespective of whether they are of equal status or, as in Tom's predicament, of unequal status.[3]

The perplexing multi-face fellow-feeling is, as shown below, at the root of debates

---

[2] In the 1962 film version, using same tile, starring Gregory Peck as "Atticus Finch," the defense attorney, Mr. Gilmer added "a white woman" when he stated: "*You* felt sorry for *her*, a white woman, you felt *sorry* for her?"

[3] What about the exchange of goods in relation to status inequality—the issue that enlivens the work of Thorstein Veblen [1934]? Is it naïve to assume that the exchange of goods among agents has the same meaning irrespective of whether they are of equal or unequal status? This is issue is the core concern of the classical labor-theory of value, especially Karl Marx's [1976, ch. 1] concept of "abstract labor" [see Khalil, 1992]. Classical theory, as well modern theory, assumes naively that exchange of goods disregard the issue of status. That is, two goods that cost the same should sell for same price in competitive markets—ignoring the role of brand names and snob appeal. This naïve assumption of economics was challenged, although on unnecessarily repugnant racial and colonialist grounds, by Thomas Carlyle in the 19th Century. As David Levy [2001] demonstrates, Carlyle dubbed economics the "dismal science" exactly because it ignores the role of status inequality in the exchange of products [see Khalil, 2007a].

concerning the different functions of the mirror neuron system (MNS) and, generally, the canonical neuron system (CNS). These systems were recently discovered in particular regions of the brain of primates and other mammals. They are usually identified as the seats of fellow-feeling expressed by an observer (called "spectator") towards the action of an actor (called "principal"). The MNS of the spectator is instigated when the spectator *understands*, i.e., empathizes with the intention of the principal. The CNS of the spectator, but also MNS to some extent especially if the spectator is human, facilitate *imitation* whose function, by definition, ignores the intention of the principal actor, i.e., does not involve the function of understanding (empathy).

Obviously, the "understanding function" and the "imitation function" of fellow-feeling differ. If we ignore the particular situation, social relations, or the relevant institution, the difference of functions is perplexing, what is called here the "mirror-neuron paradox." How could the same emotion, fellow-feeling, function as "understanding" and as "imitation"?

Amazingly, David Hume has long ago noticed the *same* contradictory functions of fellow-feeling. That is, what is called here Hume's "fellow-feeling paradox" is identical to the mirror-neuron paradox. To wit, Adam Smith tried to solve Hume's paradox. Smith developed a concept of sympathy that is remarkably identical to the "understanding function" of MNS, as few authors have recently noted [see Rustichini, 2005; Rizzolatti & Craighero, 2004a].[4] Nonetheless, Smith fails in solving the fellow-feeling paradox, viz., how could imitation coexist with understanding?

The central aim of this paper is to solve the paradox. The paper's argument proceeds along

---

[4]There is an intimate link between MNS (fellow-feeling) and interpersonal utility comparison [see Fontaine, 2001]. But this ramification of MNS is not pursued here.

the following sections:

1. How do the interpretations of the recent MNS give rise to the "mirror-neuron paradox"?
2. How is mirror-neuron paradox is nothing but the fellow-feeling paradox first noted by David Hume?
3. In response to Hume's challenge, Smith has attempted a solution. Smith has explicitly argued that the "understanding function" of fellow-feeling necessarily entails the approbation of action (propriety), which is a judgment of whether the action is optimal. This clearly sets the "understanding function" far apart from the "imitation function" of fellow-feeling. Nonetheless, Smith cannot explain the coexistence of the two functions.
4. Contrary to Smith, the "understanding function" of fellow-feeling does not entail the approbation of action (propriety). To show this, this core section distinguishes between the "rationality axis," which occasions approbation or disapprobation, from the "intentionality axis," which occasions understanding (empathy) or revulsion. The proposed "two-axis evaluation" hypothesis (TAE) promises to solve the mirror-neuron paradox, i.e., to explain the coexistence of the "understanding function" and the "imitation function."
5. This section suggests an experimental set-up to test the TAE hypothesis.
6. One major payoff of the TAE hypothesis is the analytical tool kit that allows us to distinguish among four kinds of fellow-feeling, each processed along different axes: "sympathy," "compassion," "indulgence," and "adulation."
7. The analysis of adulation is necessary for modeling status inequality and it is at the origin of

political authority and "*assabiya*"—the Arabic term that Ibn Khaldûn [1967] uses to denote tribalism, group identity, team spirit, or, in short, group solidarity. The literature on "social preferences," insofar as it fails to recognize the intentionality axis, cannot distinguish group solidarity apart from either altruism or fairness.

8. The opposite of altruism is selfishness. But the literature uses terms such as "spite" and "malevolence" interchangeably with "selfishness." To wit, insofar as economic theory does not recognize the intentionality axis, it lacks a theory of evil. A theory of evil would allow us to distinguish, first, selfishness from malevolence and, second, racial segregation from racial hatred.

1. The Mirror-Neuron System (MNS)

The discovery of the mirror neurons is largely attributed to the laboratory of Giacomo Rizzolatti [1999; 2004, in Hurley & Chater, 2005]. The amazing central feature of MNS is that it becomes instigated irrespective of whether the spectator undertakes an action, such as grasping an object of significance (cup), or the spectator watches another (called throughout "principal") undertaking this action. The MNS was first discovered in monkeys, located mainly in F5 area of the brain, but later found in dogs and humans:

> Mirror neurons are a particular class of visuomotor neurons, originally discovered in the area F5 of the monkey premotor cortex, that discharge both when the monkey does a particular action and when it observes another individual (monkey or human) doing a similar action [Rizzolatti & Craighero, 2004, p. 169].

MNS has the following general characteristics:

1. The object significance, whether grasping a cup or food, is insignificant as to whether the

spectator's MNS is instigated or not.

2. The observed subject (principal) can be close or far away from the spectator without a difference.
3. The principal can succeed and be rewarded with the action or can fail—without a difference for the discharging of MNS.
4. The spectator's MNS is instigated even when the spectator and the principal belong to different species.[5]
5. When MNS discharges, it combines the emotion triggered by the stimulus and the action in response to the stimulus. That is, there is no dichotomy between emotion and action. Throughout the paper, no distinction is made between the two. To wit, agents who feel, but to do not act, it is because the action is inhibited by another neural system that takes into consideration other factors.

For our purpose here, the most important feature, stressed by Rizzolatti & Craighero [2004, p. 170], is that MNS is based on "transitive motion," where MNS is instigated when the spectator observed action moves towards a purpose, such as a hand reaching for a cup. MNS is usually dormant when the spectator observes only "intransitive motion," i.e., action that has no goal or meaning such as the motion of hand with no cup in sight. Such meaningless, intransitive motion does not instigate MNS. But it does instigate another system, called "canonical neurons":

> There are two classes of visuomotor neurons in monkey area F5: canonical

[5] Given MNS operates across nonconspecifics, some institutions can be interpreted as inhibitions. For instance, "halal" (Islamic rule) and "Kosher" (Judaic rule) inhibit the mirror-neuron system, allowing humans to suspend fellow-feeling with animals categorized as food.

> neurons, which respond to the presentation of an object, and mirror neurons, which respond when the monkey sees object-directed action. In order to be triggered by visual stimuli, mirror neurons require an interaction between a biological effector (hand or mouth) and an object. The sight of an object alone, of an agent mimicking an action, or of an individual making intransitive (nonobject-directed) gestures are all ineffective [Rizzolatti & Craighero, 2004, p. 170].

Broadly speaking, then, MNS involves understanding, i.e., the spectator understands the intention of the principal's action. So, it must involve object-directed action. In contrast, canonical-neuron system (CNS) seems to involve imitation, i.e., where the issue of intentionality of the principal is irrelevant. The principal simply replicates the action/emotion of the principal without taking into consideration its goal or object. So, CNS apparently does not worry about whether the action is repulsive or understood. It contrast, MNS apparently is engaged when understanding is involved.

In understanding, the spectator cannot represent the emotion/action of the principal without examining as well the intention of the principal's action. Further, it might involve a judgment of whether the action is suitable or proper given the stimulus or what economists call incentive. In imitation, the spectator represents the emotion/action of the principal's without any attention to the stimulus, i.e., the action is not being examined in relation to the stimulus that occasions it.

But from the discussion in the literature, there is no clear differentiation between MNS and canonical-neuron system [see Hurley & Chater, 2005, vol. 1, ch. 1]. Rizzolatti and Craighero [2004] even argue that MNS is involved in both functions, understanding and imitation. They maintain that language acquisition is greatly based on imitation, where the spectator (child) mimics the adults (principals) without understanding.

While the two functions, viz., understanding and imitation, somewhat overlap, it is important

to distinguish them. With imitation, there is no understanding. With understanding, there is no imitation. Given that the two functions are different, how can we distinguish them? It is insufficient to trace them back to some neural substrate. Still, how could one neural substrate be invoked with respect to imitation but dormant with respect to understanding?

Interestingly, the two functions (imitation and understanding) of mirroring should prove to be the core of the paradox that David Hume, long ago, has highlighted.

2. Hume's Fellow-Feeling Paradox

Hume challenged his friend, Smith, with a paradox, called here the "fellow-feeling paradox."[6] In his 28th July 1759 letter to Smith, Hume posed the following question: Why does sympathy with someone in grief over the loss of a child usually attenuates the sense of grief, rather than leads to the reinforcement and escalation of grief?[7]

> I am told that you are preparing a new Edition [2nd edition of *Theory of Moral Sentiments*] and propose to make some Additions and Alterations, in order to obviate Objections. I shall use the Freedom to propose one, which, if it appears to be of any Weight, you may have in your Eye. I wish you had more particularly and fully prov'd, that all kinds of Sympathy are necessarily Agreeable. This is the Hinge of your System, and yet you only mention the Matter cursorily in p. 20. Now it woud [sic] appear that there is a disagreeable Sympathy, as well as an agreeable: And indeed, as the Sympathetic Passion is a reflex Image of the principal, it must partake of its Qualities, and be painful where that is so. Indeed, *when we converse with a man with whom we can entirely sympathize,* that is, where there is a warm and intimate Friendship, the cordial openness of such a Commerce overpowers the Pain of a disagreeable Sympathy, and renders the whole Movement agreeable. But in

[6] Eric Schliesser alerted me to the fellow-feeling paradox. David Levy and Sandra Peart [2004] brought the letter concerning the paradox to my attention.

[7] This phenomena, how sympathy with someone in grief, gives a sense of joy has also fascinated the mystic philosopher Edith Stein [2002] in her analysis of how the suffering associated with the Christian cross affords a sense of joy.

> ordinary Cases, this cannot have place. An ill–humord Fellow; a man tir'd and disgusted with every thing, always *ennuié;* sickly, complaining, embarass'd; such a one throws an evident Damp on Company, which I suppose wou'd be accounted for by Sympathy; and yet is disagreeable [Hume in Smith, 1977, p. 43].

To express Hume's fellow-feeling paradox,

$$E^{d}_{p} = E^{d}_{p}[E_{s}(E^{o}_{p})]$$

whereas E^{d}_{p} is the principal's *derived* emotion; E_{s} the spectator's emotion; E^{o}_{p} the principal's *original* emotion. That is, the principal's original emotion influences the spectator's, which in turn influences the principal's derived emotion.

Fellow-feeling attenuates the original emotion (what Hume calls "agreeable" sympathy) when,

$$\partial E^{d}_{p}(E_{s})/\partial E_{s} < 0$$

In contrast, fellow-feeling reinforces the original emotion (what Hume calls "disagreeable" sympathy) when,

$$\partial E^{d}_{p}(E_{s})/\partial E_{s} > 0$$

The paradox lies in the following: How can the same building block of emotion, fellow-feeling, perform two contradictory functions: "break pedal" and "accelerator pedal"?

While Adam Smith focused on the "break pedal" function, attenuation, as discussed below, most economists have not noticed it. They rather noticed the "accelerator pedal" function, reinforcement. For instance, Gary Becker's [1991; 1996] theory of social interaction is based on the "accelerator pedal" function of emotion/action. The theory shows how particular preference can

escalate into a fad [see also Karni & Schmeidler, 1990]. Interestingly, Friedrich Nietzsche condemned Christianity for the same reason. Namely, Christianity promotes "*mitleiden*" (German: *mit*=with, *leiden*=suffering). Unfortunately, the German word "*mitleiden"* is translated into "pity" rather than suffering—given that the term "pity" denotes demeaning status inequality. In any case, Nietzsche's suffering in Christianity is self-indulgence because it is self-centered and, hence, contagious via imitation, i.e., it leads to the escalation of suffering that may push people into lethargy and depression:

> Christianity is called the religion of *pity*. Pity stands in antithesis to the tonic emotions which enhance the energy of the feeling of life: it has a depressive effect. One loses force when one pities. The loss of force which life has already sustained through suffering is increased and multiplied even further by pity. Suffering itself becomes contagious through pity; sometimes it can bring about a collective loss of life and life-energy which stands in an absurd relation to the quantum of its cause (--the case of the death of the Nazarene)" [Nietzsche, 2006, p. 488].

Depression, given its contagious character, is at the root of the model of Douglas Bernheim and Oded Stark [1988] concerning what they call "altruism." They reasoned that "nice guys," i.e., altruists, might finish last because no one would want to marry them. Why? Let us say that the partner is depressed. These nice guys would express their sorrow sympathies in the sense of the "imitation function." The partners consequently would, as a result of imitation, would feel even more depressed.

3. Smith's Solution of the Paradox

The first paragraphs of *The Theory of Moral Sentiments* betray the fact that Smith [1976] took Hume's paradox very seriously. The paragraphs show Smith's major theoretical innovation: Not all

fellow-feelings are alike. What matters for sympathy is that the spectator *is* considering the incentive (stimulus) that occasions the action/emotion of the principal. The consideration of the incentive, prior to issuing sympathy, is responsible for the attenuation, break pedal function of fellow-feeling.

Some commentators have noted the fact that sympathy, for Smith, attenuates the principal's emotion [see Haakonssen, 2002, p. xiv; Levy & Peart, 2004, p. 334, n. 3]. They note that it is a paradox for Hume because Hume is operating under another concept of "sympathy," viz., as imitation that is responsible for the reinforcement, "accelerating pedal" function. As Philippe Fontaine [1997] and Robert Gordon [1995] show, David Hume defined sympathy as emotional contagion or, what is the same mechanism, projection of one's feeling onto others.[8]

So, to solve Hume's paradox, Smith simply focused on the "break pedal" nature of sympathy. For sympathy to act so, the spectator must be transporting himself into the station of the principal, and trying to enter his emotion/action by examining the cause or incentive that gave occasion to the emotion/action.

If this is the case, the act of sympathy in Smith is nothing other than the act of rational choice in modern economics. Here, the spectator is examining, by putting himself in the shoes of the principal, whether the principal is reacting efficiently, i.e., proportionally to the incentive. So, the issue of rational action/emotion is at the heart of the analysis of sympathy.

To elaborate, to solve Hume's paradox, Smith redefines and narrows the meaning of the term

[8] Fontaine seems aware of the problem of defining sympathy as emotional contagion. In contrast, Robert Gordon is uncritical of the definition. He shows how modern neuroscience questions Hume's assertion that cognition intervenes between perception and emotions.

"sympathy." Sympathy is not the contagious emotion suggested by imitation. Rather, it is about understanding. As such, one should not be perplexed, as the case with Hume, when sympathy attenuates the original emotion.

To wit, Smith's concept of sympathy-as-understanding corresponds well with the "understanding function" of MNS. Namely, in its function as understanding, the spectator's MNS becomes instigated only when the spectator observes that the principal is involved in transitive action, i.e., objected-directed action or action in relation to stimulus (incentive). So, the spectator does not simply imitate the emotion/action of the principal. The spectator can only replicate the principal's emotions if such emotions are understood, i.e., how is the action related to the stimulus.

But why should the "understanding function" leads to the attenuation of the original emotion? And why such understanding gives another layer of satisfaction? These two phenomena perplexed Hume. Smith provides a single answer that remarkably explains both phenomena.

3.1 Attenuation of the Principal's Emotion

Concerning the first phenomenon, attenuation of original emotion, the spectator who is trying to understand the emotions of the principal is not any spectator. Rather, he or she is an "impartial spectator." Smith's concept of the "impartial spectator" and the mechanism that gives rise to what he calls "propriety" is involved [Khalil, 1990, 2006]. Stated briefly, as an impartial agent, the spectator is, by definition, a judge of whether the principal's action/emotion is proportional to the stimulus (incentive). That is, the judge has to determine if the action is optimal given the incentive. When the principal also acts as the judge, then the impartial spectator resides within the principal, as

second self, and called throughout the "judge within." The judge within, to note very briefly, is *not the internalization* of social norms à la functionalist sociological theory or microsociological theory à la George Herbert Mead [see Khalil, 1990]. Rather, for Smith, the judge within is simply the principal looking after the utility of the current self as well as the utility of the future self—i.e., examining the demands of each self from a distance that is occupied by the impartial spectator as well. So, there is a hall of mirrors in Smith's analysis—where there is a correspondence between the judge within and the judge without (impartial spectator). But it is important not to get lost in such a hall and keep our focus on the "original copy." For Smith, and in this paper, the "original copy" is the judge within, which is usually *externalized* in case our agent is not a Robinson Crusoe. The original copy is not the judge without or what agents think is the judge without, as implied in the analysis of Jean-Pierre Dupuy [2004].

In this analysis, the principal, to be rational, takes action in light of the choice determined by the judge within. This interpretation of Smith's theory makes it a theory about "self-command," which is one of the main virtues of Smith's book. When the principal exercises self-command, the principal is accommodating the needs of the current self which, given the scarcity of resources, competes with the needs of future self. And the principle, or judge within, is capable of doing such an accommodation by examining the needs of the current self from a distance, which Smith calls metaphorically the "impartial spectator." That is, the impartial spectator is nothing but a metaphor for non-myopic decision making, where the utility of the future self must count. So, there is no fissure between the judgment of the caring, impartial spectator and the judgment of the self concerning its own welfare--a fissure that is supposedly deep and need to be bridged according to

Stephen Darwall [2002, 2006].

So, Smith's theory of sympathy is ultimately about rational intertemporal allocation when time inconsistency (temptation) is a problem. Smith's theory anticipates the dual-self model of intertemporal choice that is gaining attention not only in economics [Thaler & Shefrin, 1981; Fudenberg & Levine, 2006] but also in biology with regard to intrapersonal conflicts [e.g., Haig, 1993, 2003; Burt & Trivers, 2006]. This might come as a disappointment for the new scholarship on Smith, such as Deirdre McCloskey's [2006], which promotes Smith as an alternative to narrow standard theory of choice [e.g., Gintis *et al.,* 2005]. Smith's concepts of sympathy and self-command are ultimately about optimal choice.

However, Smith's theory is not this simple. It provides a rich account of the mechanics of self-command, i.e., how exactly does the self enforce time consistency and succeed in fighting temptations? Here, Smith divides the agent into principal and impartial spectator, which is expressed externally as "doer" and "judge," respectively [see Khalil, 1990]. If the principal (doer) becomes very agitated as a result of a simple failure, and surrenders to anger, the impartial spectator (judge) would not sympathize with the principal. This means that the impartial spectator or, for short, the spectator cannot approve the principal's action/emotion. For the principal to win the approbation of the spectator, the principal must take residence in the spectator's station, i.e., look at his current anger from a distance. Such an examination would allow him to see that if he acts with anger, he might hurt future self. So, a judge has to restrain current self so that the future self is not hurt. But how does this exactly work? For Smith, the current self seeks the sympathy of the judge. The judge, acting, as a spectator, cannot provide sympathy, i.e., approval about the efficiency of the

action, if the pitch of action/emotion of the principal is too high or disproportional to the cause (incentive). The principal, hence, must lower the pitch of emotion/action to win the approbation, i.e., sympathy, of the spectator. So, the act of sympathy can be interpreted as nothing but the fact that the judge is taking into consideration the interest of future self as well.

If the principal lowers the pitch of emotion/action, it would be easier for the spectator to travel and enter, i.e., sympathize, the station of the principal. As long as the principal is too angry or too joyful relative to the incentive, the impartial spectator simply cannot understand the emotion/action of the principal, i.e., approve. So, for Smith, the spectator's understanding (empathy) automatically entails approbation, i.e., judgment of propriety. Smith's notion of "sympathy" is nothing but the collapse of empathy (understanding), on one hand, and propriety (efficiency of action), on the other. So, sympathy is nothing but the conclusion that the principal has acted in his interest and effectively (optimally), i.e., he has not only chosen welfare-enhancing path, but he has also chosen the optimal path, given the incentive. For Smith, sympathy occurs only when the action or choice is both welfare-enhancing and optimal.

3.2 Another Layer of Satisfaction

Concerning the second phenomenon, there is another layer of emotion that accompanies Hume's "agreeable" sympathy (understanding). This layer is absent in "disagreeable sympathy" (imitation) which leads to the escalation of fellow-feeling. Namely, given that the spectator's understanding entails the also approbation of the principal's emotion or action, such sympathy affords "another source of satisfaction" [Smith, 1976, p. 14]. This additional layer of emotion is always positive—

irrespective of whether the original emotion was grief over bad news or joy over good news. Such second layer of emotion is self-satisfaction or self-congratulation that one has exercised what Smith calls "self-command" with regard to the original emotion—whether grief or joy. The principal, upon succeeding in calibrating the action in proportion to the stimulus (incentive), whether grief or joy, the principal has acted with propriety, i.e., optimally.

Consequently, the principal is infused with a sense of self-congratulation, accomplishment, integrity, or what can be called in general "symbolic utility" [Khalil, 2000a]. This is considered a second layer of emotion because it cannot exist independently of taking the proper or optimal action and, hence, it is called "symbolic." Symbolic satisfaction arises also when one succeeds in exercising self-command over the appetites when one encounters a tray desert, when one has a commitment not to indulge. Likewise, when one controls his joy over good news, one derives utility from not celebrating in a careless fashion and also derives a sense of integrity for being so prudent. The same occurs when the original emotion is grief. When one controls his grief over bad news, one derives utility from not giving in to anger and also derives a sense of integrity for being so prudent. While the success of the resisting temptation affords a greater utility, given that (discounted) health is preferred to momentary pleasure, it also affords the second layer of satisfaction, namely, the symbolic effect.

Smith, in fact, directly criticizes Hume, the "ingeneious and agreeable author," for postulating that there is only one source of satisfaction, viz., utility. For Smith, Hume fails to recognize integrity, the self-satisfaction arising fro acting with propriety that accompanies "approbation":

> The same ingenious and agreeable author who first explained why utility pleases, has been so struck with this view of things, as to resolve our whole approbation of virtue into a perception of this species of beauty which results from the appearance of utility. No qualities of the mind, he observes, are approved of as virtuous, but such as are useful or agreeable either to the person himself or to others; and no qualities are disapproved of as vicious but such as have a contrary tendency. And Nature, indeed, seems to have so happily adjusted our sentiments of approbation and disapprobation, to the conveniency both of the individual and of the society, that after the strictest examination it will be found, I believe, that this is universally the case. But still I affirm, that it is not the view of this utility or hurtfulness which is either the first or principal source of our approbation and disapprobation. These sentiments are no doubt enhanced and enlivened by the perception of the beauty or deformity which results from this utility or hurtfulness. But still, I say, they are originally and essentially different from this perception [Smith, 1976, p. 209 (TMS IV.2.3)].

That is, for Smith, the source of satisfaction related to approbation, what is called here the second layer of emotion, or the sense of integrity, is "originally and essentially different from this [utility] perception."

The second layer of emotion, the self-satisfaction arising with the sense of integrity or approbation, goes to show how Hume's "agreeable" sympathy, i.e., in its "understanding function," is possible. First, the agreeable sympathy can explain the attenuation of the original emotion/action and second, via approbation, can account for the sense of integrity that is always positive—irrespective of whether the original emotion is joy or grief. So, agreeable sympathy with a grieving person leads to the attenuation of grief on two counts: first, through the lowering of the original pitch and, second, upon succeeding in lowering the original pitch, the grieving person experiences self-satisfaction for being rational. (But, note, if the agreeable sympathy is with a joyful principal, it would attenuate the pitch of joy, on one hand, and lead to self-satisfaction for being optimal, on the other hand.)

3.3 Sympathy with the Dead and Insane

Smith further defends his notion of sympathy, a notion that can account for the attenuation of the principal's emotion, by pointing out that his notion can make sense of how sympathy with the insane and dead is possible [Smith, 1976, p. 54]. This would be perplexing—in fact Sugden [2002] finds it incoherent—if sympathy is merely Hume's "disagreeable sympathy" that escalates original emotions.

As argued by Smith in the paragraph that immediately precedes his discussions of sympathy with the insane and dead, sympathy in general is rather the outcome of transporting oneself to the station of the other, and feeling what the other would feel:

> Sympathy, therefore, does not arise so much from the view of the passion, as from that of the situation which excites it. We sometimes feel for another, a passion of which he himself seems to be altogether incapable; because, when we put ourselves in his case, that passion arises in our breast from the imagination, though it does not in his from the reality. We blush for the impudence and rudeness of another, though he himself appears to have no sense of the impropriety of his own behaviour; because we cannot help feeling with what confusion we ourselves should be covered, had we behaved in so absurd a manner [Smith, 1976, p. 54].

So, when one feels sorry for the dead or insane, one is expressing what they are missing if they are alive or healthy. In the case of the dead, the sorrow is a negative function of age and a positive function of creativity or accomplishments. The agent cannot reach such judgment if he was merely imitating their feelings—which do not exist. Also, he cannot reach such judgment via projection—because it is obvious that the dead or insane are no longer viable recipients of such projections. Further, he cannot reach such judgment via self-centered indulgence of remembrance because the agent would not have been dead or insane in the past.

3.4 Why Smith's Solution Fails

Smith has succeeded in setting Hume's agreeable sympathy, which attenuates the original emotion, far apart from Hume's disagreeable one, i.e., the one that simply escalates the original emotion. However, has Smith succeeded in solving Hume's fellow-feeling paradox, i.e., explaining the coexistence of the two sympathies? As quoted above, Hume has confronted Smith with the following: "I wish you had more particularly and fully prov'd, that all kinds of Sympathy are necessarily Agreeable. … Now it woud [sic] appear that there is a disagreeable Sympathy, as well as an agreeable."

In short, Smith has failed to show how the same primitive emotion, fellow-feeling, can also give origin to Hume's disagreeable sympathy. To put it in the terms of the mirror-neuron paradox, Smith showed the roots of the "understanding function" of fellow-feeling, which is behind attenuation of emotion. Smith does not reconcile the "understanding function" with the "imitation function" of fellow-feeling, which is behind the reinforcement of emotion.

This does not mean that Smith was unaware or ignorant of the "imitation function" of fellow feeling. In fact, Smith recognizes it when he discusses, e.g., the pleasure of company when people read a book together as opposed to reading it alone [Smith, 1976, p. 14]. As Martin Hollis [1998] notes, it is usually pleasurable to converse with people who had similar experiences. When one reads a book, watches a film, purchases a new automobile, or dines at a restaurant, it would be more pleasurable to converse with others who had undergone the same experience. Such conversation enhances the marginal utility as a result of the escalation effect. Gary Becker [1991, 1996; Becker & Murphy, 1993] argues that such social dynamics of consumption is

responsible for fads.

To recognize the "imitation function" of fellow-feeling, which occasions escalation behind fads, is one matter. It is another matter to show how the same fellow-feeling can occasion escalation in one case and attenuation in another. Smith failed to show such double function of fellow-feeling. Thus, he failed to resolve the mirror-neuron of Hume.

Robert Sugden [2002], in his interpretation of the paradox, also ignores the "imitation function" of fellow-feeling. In fact, when Sugden confronts the issue of how can sympathy with a grieving principal leads the principal to feel self-satisfaction (integrity), Sugden does not ground integrity in rationality, while Smith grounds integrity in rationality. Sugden rather invokes some moral principle that accompanies the sympathy with grieving person.

In contrast, this paper adheres, at least on this point, with Smith. Namely, we do not need, à la Sugden and others, need moral principle outside of rationality in order to account for the second layer of satisfaction—i.e., integrity. After all, for Smith, one would experience integrity if one is already acting with propriety, i.e., rationally. However, this paper must eventually disagree with Smith, if it aims to solve the mirror-neuron paradox.

4. The Two-Axis Evaluation Hypothesis (TAE)

The mirror-neuron paradox amounts to how could the same primitive, fellow-feeling, give rise to imitation (which involves the escalation of original emotion) and understanding (which involves the attenuation of original emotion). To show this, we need to take issue with a major thesis of Smith. Namely, Smith argues that the "understanding function" of fellow-feeling necessarily entails

approbation of action (propriety). Smith needed to make such an assumption to account for why sympathy with a grieving parent, involving understanding, does not escalate the original grief.

But does every "understanding" involve approbation? From casual empiricism, one may understand the action of Israel's little 2006 summer war, in which it killed over a thousand Lebanese, over 90% civilian, in retaliation for Hezbollah's earlier action that involved kidnapping two Israeli soldiers. But does such understanding entail approval of efficiency, i.e., the action was calibrated to the cause? Likewise, one may understand President Truman's decision to drop the nuclear bombs on two Japanese cities or the US-UK invasion of Iraq, but does it entail approbation? As discussed above, approbation means that the action taken is proper, i.e., proportional to the cause, rather than surrendering to anger and reckless behavior in response to enticing opportunity. One may understand that one, under the temptation of superior power or the temptation of a desert tray, succumbs to weakness of will and acts.

If we accept the casual empiricism, approbation concerning the propriety of action is simply a question about the optimality of emotion/action, while understanding is related to whether the intention of the actor. For instance, an agent may have a commitment to restrain himself even if he has momentary military superiority or instantaneous confrontation with a desert tray. And to act contrary to either commitment, i.e., react proportionally to the stimulus, makes us judge the action as improper or what economists call "inefficient." But such judgment does not entail that we failed to understand the principal's utility and constraint functions. If we judge *ex ante* that Truman acted improperly or inefficiently when he approved the use of nuclear weapons, it does not mean we do not understand why he did so. Truman's intention is to enhance welfare by saving the lives of

American soldiers, to bring a speedy conclusion to the conflict that may save more Japanese lives, and to secure unconditional Japanese surrender.

In this light, the act of understanding (empathy) need not entail approbation of propriety, i.e., judgment concerning the rationality of the act. And *vice versa*, the judgment concerning rationality does not entail empathy. For instance, we can be impressed with the efficiency of a serial killer, the Nazi Holocaust organizer, or a cult leader. But this does not entail that we understand, in the sense of empathize, with the intention of the agent.

So, we should question Smith's conflation of understanding with approbation of propriety and *vice versa*. But the rejection of the conflation cannot be exclusively based on, or motivated by, casual empiricism. What is exactly the payoff of rejecting Smith's concept of sympathy that insists on the conflation of understanding with approbation propriety? The payoff, as already mentioned, is nothing but the solution of the mirror-neuron paradox which Hume long ago noted.

So, contrary to Smith, this paper conjectures that we understand along an axis that is orthogonal to the axis concerning the evaluation of efficiency. Let us call the axis that may give rise to approbation of optimality the "rationality axis," while call the axis that may give rise to understanding the "intentionality axis." The rationality axis occasions the familiar judgment of action as either proper (efficient) or improper (suboptimal). The intentionality axis occasions either revulsion or empathy.

The central innovation of this paper lies, first, in identifying the intentionality axis and, second, in separating the intentionality axis from the rationality axis. First, concerning identification, the term "empathy" is opposed to "revulsion"—and this is a crucial juxtaposition.

Many authors have concluded that if we define "sympathy" along with Smith, i.e., as moral approval of the action, "empathy" is "understanding" in the sense of comprehending, as for example one comprehends the trajectory of rocket as depending on momentum energy, friction, gravity, and so on. These authors actually confuse *understanding* with *comprehension* [e.g., Binmore, 1998; Harsanyi, 1977].[9] Comprehension entails scientific examination of why hurricanes, genocides, and serial killing take place. In contrast, understanding or, interchangeably, empathy, involves rather an evaluation. But such evaluation is not about rationality—an issue which might have caused the conflation of understanding with comprehension. The evaluation implied by empathy is rather about the evaluation of the *intention* of the actor. So, the term "empathy" is used here in the same sense as when it was coined.[10] Namely, empathy means that one understands the intention of an action of, e.g., an angry woman catching her husband cheating on her—while not passing a rationality judgment on whether her action is proper or not. The opposite, revulsion, means that one cannot

---

[9] Ken Binmore uses the term "empathy" in the sense of comprehension when he describes how a gunfighter wants to know the position of an opponent:

> Adam sympathizes with Eve when he so identifies with her aims that her welfare appears as an argument in his utility function. ... The extreme example is the love a mother has for her baby. Adam empathizes with Eve when he puts himself in her position to see things from her point of view. Empathy is not the same as sympathy because Adam can identify with Eve without caring for her at all. For example, a gunfighter may use his empathetic powers to predict an opponent's next move without losing the urge to kill him [Binmore, 1998, p. 12].

Also Harsanyi [1977] uses the term "empathy" in the sense of comprehension and assessment of position of others (opponents or loved ones). Harsanyi distinguishes empathy from "subjective preferences" or what Binmore [1994, 1998] and Amartya Sen [1977] call "sympathy." Psychologists, such as Michael Basch [1983], also use the term "empathy" in the sense of comprehension.

[10] According to Gladstein [1984, p. 40; see also Gladstein, 1987], the term "empathy" was coined in 1909 as a translation of the German *einfühlung* (from ein "in" + fühlung "feeling"). The German word, popularized by Lipps, was coined in 1858 by German philosopher Rudolf Lotze

understand the intention of an action of, e.g., a serial killer—while, again, not passing a rationality judgment.

Second, concerning separating the intentionality axis from rationality axis, the separation is imperative if we want to model revulsion or disgust. The emotions of revulsion and disgust are complex [see Miller, 1997; Rozin *et al.,* 2000]. As defined here, revulsion or disgust is the feeling that arises when one determines, rightly or wrongly, that the item of consumption is actually detrimental to one's wellbeing. In this definition, revulsion differs from squeamishness. While revulsion arises from evaluation that involves wellbeing, squeamishness does not involve such evaluation. It is based on self-centered memories and associative feelings and, hence, as discussed below, it is a form of indulgence.

To illustrate revulsion, a spectator may determine that eating particular meat, such as the meat of snakes or alligators, is revolting. Such revulsion would be based on the evaluation, at least at the gut feeling level, that it is detrimental to one's wellbeing. Our spectator may even experience nausea and sickness in the stomach when he sees a principal eating the item under focus.

The problem is the following: Let us say that our spectator, who is totally repulsed by the meat of reptiles, volunteers and eats the item for no apparent or hidden compensation. How can we model such agent who undertakes actions that they fully know to be destructive or detrimental of wellbeing? To argue that the spectator must have changed his or her mind, and now prefers the item, amounts to *ad hoc* reasoning as if preferences are unstable [Stigler & Becker, 1977].

This problem actually puts in a new light what Robert Frank [2006, p. 231] calls the

---

(1817-81) from the Greek *empatheia* "passion," from en- "in" + pathos "feeling."

"`crankcase oil' problem." The problem is based on George Stigler's famous quip: How should we model a person who drinks crankcase oil from his automobile while fully knowing that it is neither medicinal nor tasty, but it is rather detrimental to wellbeing? If we assume that the person simply likes the crankcase oil, it would violate the principle of stable preferences. We simply cannot move item Z from the category of "garbage" to "goods," assume that preferences have changed, simply because we now observe the spectator under focus consuming Z.

If we maintain the standard position, viz., the rationality axis is the only axis of evaluation, we would not be able to explain revolting or destructive behavior as illustrated in the crankcase oil problem. The drinking of crankcase oil, or having revolting intention to lower wellbeing of the self or others, is not an issue about prices and budget constraints, where the rationality axis would be relevant to make optimum resource allocation. It is rather an issue about survival or no-survival, which does not involve a question about allocation of resources.

To account for revulsion, there must be an axis of evaluation, called here the intentionality axis, which cannot be reduced to the issue of allocation of resources, i.e., the rationality axis. The orthogonality of the two axes is the core idea of the proposed "two-axis evaluation" hypothesis (TAE).

To wit, TAE makes sense of the casual empiricism mentioned above. The rationality axis asks whether the serial killer acted efficiently or inefficiently such as succumbing to opportunities that were *ex ante* clear to be suboptimal. Likewise, we can argue that Truman succumbed to myopic benefits when h used the atomic bomb—even when measured in terms of saving American lives in the longer term. In contrast, the intentionality question may find that Truman's intention is

understandable, i.e., one can empathize with it given that it mainly was about the protection of American lives. But one may not understand (empathize) with Truman if his motive was hatred, vengeance, and spite. Likewise, one would experience revulsion, not empathy, with the intention of the serial killer—even if he finds his methods to be efficient and well-calculated.

The TAE hypothesis, as shown below, can resolve the mirror-neuron paradox because fellow-feeling, or mirroring, is processed along the two axes not only when they are engaged, but also processed when they are disengaged. It is difficult to think how either axis can be suspended or disengaged. To start with, let us map the structure of possibilities: The rationality axis can be totally suspended while the intentionality axis is engaged, and *vice versa*. Or both axes are suspended or both are engaged. We have four possible combinations, as Figure 1 shows.

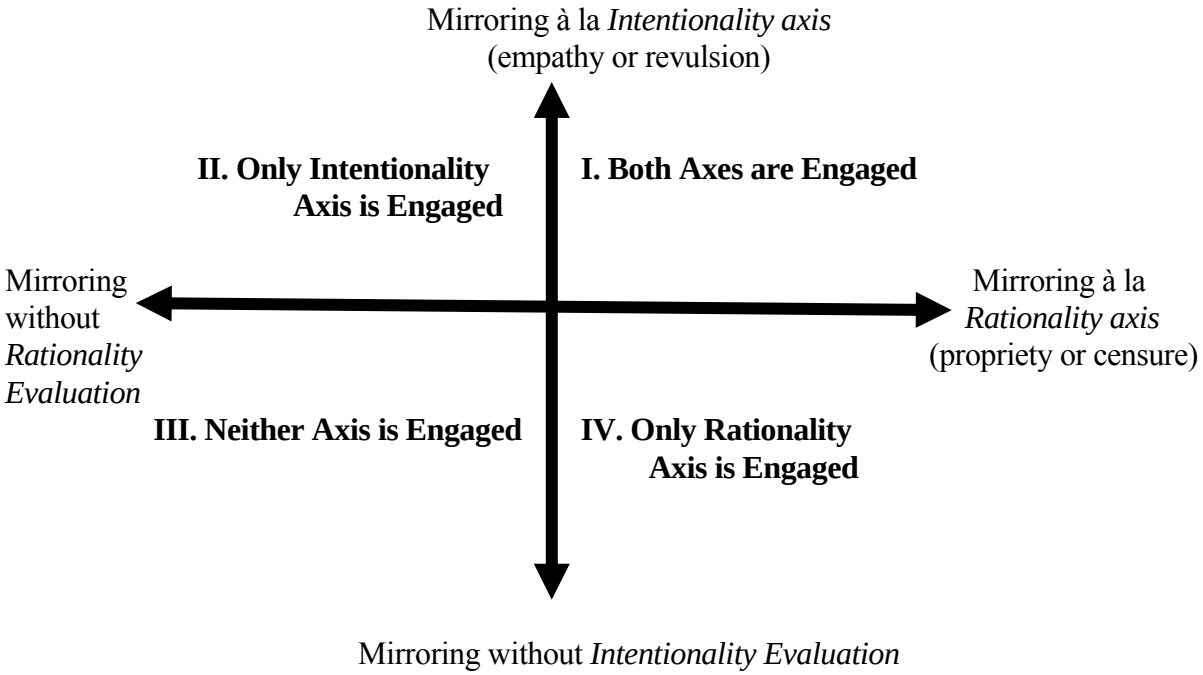

Figure 1: The Two-Axis Evaluation Hypothesis (TAE)

Quadrant I shows the combination when both axes are engaged. Quadrant II demonstrates the combination when only the intentionality axis is engaged. Quadrant III displays the combination when neither axis is engaged. Quadrant IV exhibits the combination when only the rationality axis is engaged.

For an axis to be engaged is simply, as shown above, is to ask the relevant question. If the rationality axis is engaged, one asks: is the action/emotion efficient or suboptimal? If the intentionality axis is engaged, one asks: can one empathize with the action or is it revolting? But what does it mean to have an axis disengage? For the rationality axis to be disengaged, one does not judge whether it is rational or not. The fellow-feeling or mirroring takes place without such assessment. For instance, if a serial killer commits a stupid mistake and gets caught, one may suspend the rationality axis, and simply engage only the intentionality axis (quadrant II): Is the

action of the serial killer understandable? Likewise, if a man drives fast because he is late to an appointment, and hits a crowd of people and kills a dozen of them, one may suspend the rationality axis and only engage the intentionality axis (quadrant II): is the action of the young man understandable?

On the other hand, to disengage the intentionality axis is harder to analyze. When one asks whether an action understandable—i.e., can one empathize with the agent—one is examining the principal's behavior in relation to the principal's intention. Note, we are not examining the behavior in relation to incentive—which would be a question along the rationality axis. For instance, the serial killer might have killed in total, before being caught, a half dozen people. But his intention would be examined differently from the driver who killed a dozen out of recklessness. Now, what if the intention is not considered at all? Here, the spectator processes the fellow-feeling without consideration of what motivated the principal. The spectator only senses the action without the object. But such an observation, if it registers emotion in the spectator, the emotion is evoked because of remembrance of one's own past experience. So, the emotion of the spectator has little to do with the situation. The situation is not even the subject of understanding or no-understanding. Rather, he spectator, involved in his own station or circumstances, uses the stimulus or observed action, to re-call how he would feel if the observed event happened to him.

The disengagement of the intentionality axis actually informs ego-centric theories of altruism stretching from Thomas Hobbes to Gary Becker [see Khalil, 2001, 2002b, 2004]. These theories, known also as "warm-glow" theories, the spectator/benefactor contributes to the wellbeing of the principal only insofar the excitement or utility of the principal excites, in reflection, the excitement

of the spectator/benefactor. Here, the benefactor does not care about the intention of the principal. The benefactor is only interested in how the excitement of the principal enhances his own utility.

Such a view of altruism does not distinguish between altruism and social interaction behind the rise of fads and escalation of fashion. Gary Becker [1996] lumps both phenomena almost under the same model of social interaction.

To wit, as alluded above, fads resemble the escalation of original feeling, what Hume calls "disagreeable sympathy." Here, the original emotion is amplified, and original action is extended, as others imitate the principal's action. The principal starts to reap greater marginal utility as others enact the same fashion or become in-synch with his mood. In such a situation, others imitate the principal without attention to his intention.

So, escalation of original emotion takes place when the intentionality axis is disengaged. Such escalation need not involve judgment of propriety. In Hume's example, quoted above, a merry person makes other merry, via contagion, where others do not pass judgment on the rationality of the mood. To wit, to ensure the contagious aspect of fads or moods, agents do not invoke the rationality axis.

So, the primitive fellow-feeling gives rise to escalation when the two axes are disengaged, which is depicted as quadrant III. The same primitive can give rise to the attenuation of emotion if the two axes are engaged, which is demonstrated as quadrant I. In quadrant I, even if the act is revolting—such as genocide or mass killing motivated by hate—it can still be judged according to the rationality axis. While one cannot empathize with such an act, one can still judge its efficiency. And such judgment of efficiency entails that the serial killer must not take short-cuts or given in to

excitement and anger, if he does not want to be caught.

While Smith's concept of sympathy can also, as shown earlier, explain attenuation of original fellow-feeling, it is limited in scope. It cannot explain attenuation in cases when understanding is impossible, such as in serial killing, while rationality is possible. Smith's analysis, given its conflation of understanding with propriety, lacked the analytical tools to account for wider phenomena of propriety when understanding is lacking.

Of more importance, given Smith's conflation of the two axes into one, and not realizing the consequences of suspending approbation, Smith's analytical tools cannot capture the four quadrants just discussed. Therefore, Smith's analysis of fellow-feeling cannot explain how the same primitive can lead to escalation of original emotion, and not only to its attenuation.

So, the proposed TAE hypothesis solves the mirror-neuron paradox. The same primitive, fellow-feeling, can lead to the attenuation or escalation of original fellow-feeling. This depends on whether both axes are engaged, which would lead to attenuation, or whether both axes are disengaged, which would lead to escalation.

Furthermore, the TAE hypothesis sheds brighter light on the two functions of MNS and CNS discussed earlier, viz., the "understanding function" and the "imitation function" of mirroring. When the intentionality axis is engaged, the "understanding function" or, in case of revulsion, disgust, is operative. When the same axis is disengaged, there is neither understanding nor disgust. The judgment concerning intentionality is totally shelved or frozen. In such case, the "imitation function" is operative. So, the two functions are not incompatible. The functions diverge simply because the primitive fellow-feeling is processed along different institution or different part to the

intentionality axis.

5. Testing the TAE Hypothesis

The task is, first, to test the existence of each axis and, second, to show that hey exist independently of each other.

5.1 Testing the Rationality Axis

To test the rationality axis, we can set up the following benchmark:

1. Spectators observe principals who are stimulated by incentives of different intensity (winning 1 banana to a box of fruits).
2. Records are kept of the action/emotion of principals and the corresponding spectators' MNS.
3. Principals are aware of the fact that they are being observed, but do not know the nature of the experiment.

As for the treatment,

1. Repeat steps #1-2 above
2. Principals are aware of the nature of the experiment, and their reactions are no longer of their choice. Rather their reactions are selected for them by the experimenter so that they widely differ from the benchmark case. As for the spectators, they are not informed that the reactions of the principals are manipulated.

The TAE hypothesis predicts the following. As for the intentionality axis, the spectators' canonical-neuron system (CNS) is irrelevant: it should be the same in the benchmark as in the

treatment. In both cases, there is an approval of the intentionality of principals' action since the fruits are seen to be conducive to wellbeing. The focus here is rather on MNS. If it is engaged, the spectators' MNS should behave differently in the treatment case. It should reflect impropriety. If it is not engaged, the spectators' MNS should not register any activity.

5.2 Testing the Intentionality Axis

To test the intentionality axis, it is more problematic because the wellbeing of principals cannot be harmed. Nonetheless, the harm can be measured without actually inflicting harm on the principals as shown in the treatment.

Let us start with the following benchmark:

1. Spectators observe principals who are eating "culturally understood" desert (such as most fancy ice cream with strawberry topping).
2. Records are kept of the action/emotion of principals and the corresponding spectators' MNS.
3. Principals are aware of the fact that they are being observed, but do not know the nature of the experiment.

As for the treatment,

1. Spectators observe principals who are eating "culturally disgusting" desert that is clearly knowable to the spectators (e.g., fancy ice cream with chopped liver topping).
2. Records are kept of the action/emotion of principals and the corresponding spectators' MNS.
3. Principals are aware of the nature of the experiment, and their reactions are no longer of their choice. Rather their reactions are selected for them by the experimenter so that they exhibit

the usual emotions/excitement as if they are eating "culturally understood" desert. As for the spectators, they are not informed that the reactions of the principals are manipulated.

The TAE hypothesis predicts the following. As for the rationality axis, the spectators' MNS should be the same in the benchmark as in the treatment. In both cases, there is an approval of the propriety of the action of the principals. The focus here is rather on the CNS. If it is engaged, the spectators' canonical-neuron system would behave differently in the treatment case. It should reflect revulsion or absence of empathy because the food is judged as a hindrance to wellbeing. It is similar to an act of hurting one's own body since revulsion arises from the belief that the action reduces even momentary wellbeing. If CNS is not engaged, the spectators' CNS should experience same excitement in the treatment as in the benchmark. The spectators' CNS would imitate the apparent excitement of the principals.

6. Four Kinds of Fellow-Feeling

Even if testing corroborates the TAE hypothesis, what is the payoff? One payoff is the analytical matrix needed to differentiate different kinds of fellow-feeling, including the pity that surfaced in the trial of Tom in *To Kill a Mockingbird.* Figure 2

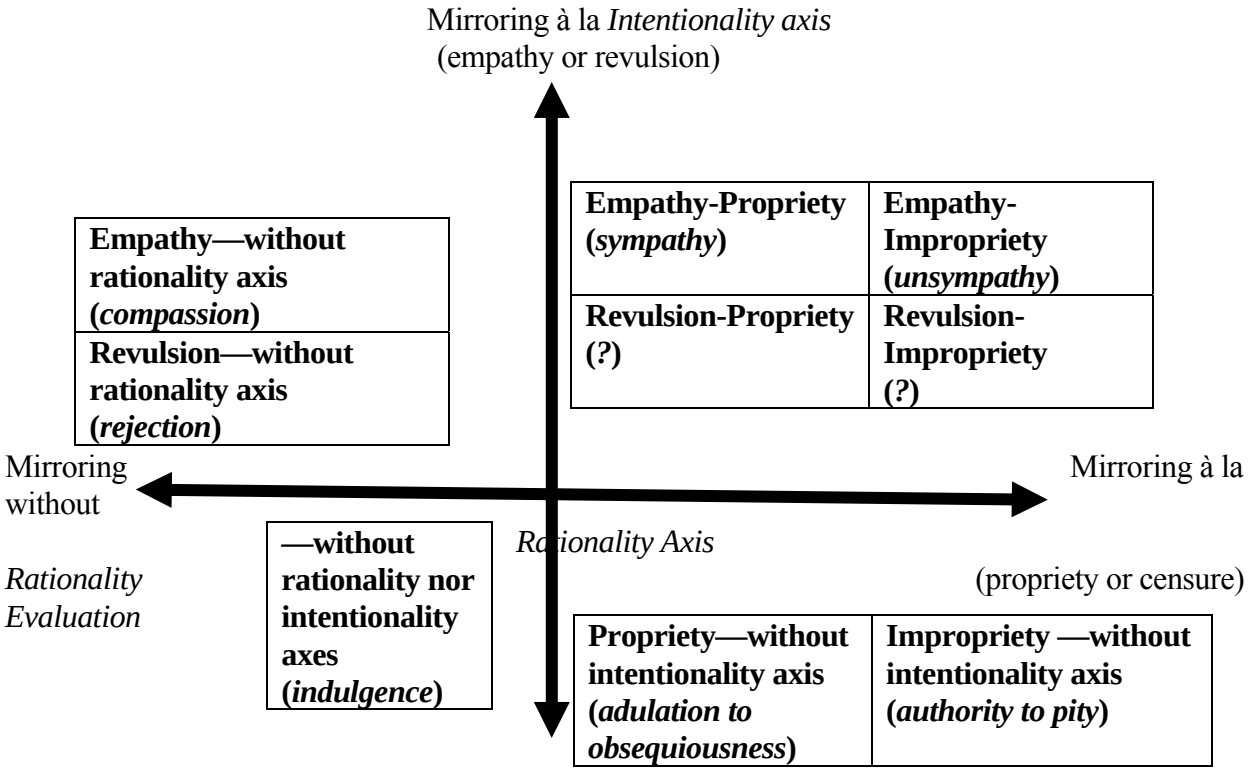

Figure 2: Four Kinds of Fellow-Feeling

reproduces the four quadrants with all possible fellow-feeling emotions that can arise when the axes are engaged or disengaged. This section selects a sample of these fellow-feelings, one from each quadrant. Namely, this section shows how the TAE hypothesis allows us to distinguish "sympathy" (quadrant I), from "compassion" (quadrant II), "indulgence" (quadrant III), and "adulation" (quadrant IV).

The choice of these terms has been difficult because the literature, even in psychology [see Lewis and Jeannette M. Haviland-Jones, 2000], has used the terms in a confusing manner. The terminological mess is understandable given that there is no theory about which scholars have agreed to be the relevant one to explain the decision making process underpinning the many faces of

fellow-feeling emotions. It is not possible to provide here even a short survey of the literature on fellow-feeling. But Figure 3 tries to give a bird's eye view of what terms

Term	Equivalent Terms used by others
"Sympathy"	"Sympathy": Smith [1976] and Scheler [1954] "Fellow Feeling": Smith [1976]; Scheler [1954] "Agreeable Sympathy": Hume [in Smith 1977]
"Indulgence"	"Pity": Nietzsche [2006] "Empathy": Lipps [1960]; Scheler [1954]; Stein [1970]; Heidegger [1962] "Sympathy-as-squeamishness": Sen [1977] "Subjective Preferences": Harsanyi [1977] "Sympathy": Binmore [1994, 1998] "Disagreeable Sympathy": Hume [in Smith 1977]
"Compassion"	"Extended Sympathy": Arrow "Empathy": Harsanyi [1977] "Empathy": Binmore [1994, 1998] "Christian Love": Stein [1970] "Mercy": Stein [1970] "Self-Love": Smith [1976]
"Adulation" ("Pity")	"Imaginative Sympathy": Smith [1976] ("Vanity"; "Pride": Smith [1976])

Figure 3: The Terminological Jumble

economists and others have used and how they correspond to the four terms differentiated here.[11]

6.1 Sympathy

Sympathy is defined as a particular fellow-feeling that may or may not arise when the rationality axis and the intentionality axis are engaged. The spectator, residing in quadrant I, expresses

[11] Fontaine [1997] also attempts to clarify the terminological mess. Fontaine contrasts "sympathy," "empathy," and what he called "partial empathy." But these categories are not broad enough to capture what is called here "ego-centricism" and "recognition." Fontaine's scope is more limited than here because his main focus is on comprehension, what Harsanyi/Binmore call "empathy." Fontaine wanted to stay within the economics literature, whose focus is to explain how people understand the constraint budget and utility of each other as they bargain in the market or maximize social welfare function. Fontaine is neither interested in ego-centricism nor in sympathy with regard to moral judgment, which are the focus here.

sympathy only when, first, approves of the intention of the principal and, second, approves of the propriety of the action. That is, along the intentionality axis, the spectator empathizes with the principal if the principal is working to enhance utility, whether his utility or the utility of a loved one. Along the rationality axis, the spectator finds the action proper if the principal acts with restraint, i.e, does not surrender to temptations or myopic emotions.

So, while sympathy entails empathy, empathy may or may not involve sympathy. While the spectator may empathize, the spectator may not sympathize if he finds the principal's action to be suboptimal. For instance, the principal could succumb to the temptation and act suboptimally in favor current self over the interest of a worthy other or over the interest of a future self. In both cases, the principal is deemed to be "selfish." Such selfishness, nonetheless, is motivated by the attempt to improve the welfare, although myopically, of the current self. So, the spectator would be empathic with the selfish principal, but he would be "unsympathetic."

Note, one should not confuse "unsympathy" with the two other possibilities, which I failed to find a proper term for them, in quadrant I. In these two other possibilities, the spectator finds the intention of the principal revolting. The spectator finds it revolting when the principal is motivated by vengeance or malice. Malice, as defined here, is an action whose sole motive is the reduction of welfare of a person just because the principal would enjoy it—i.e., not because it necessarily increase the welfare of another person. The action is judged as malice—while it can be found to be proper or improper.

As stated earlier, Smith's theory can only account for sympathy or unsympathy as defined

here. For him, it is sympathy only if the action is also understandable. It is unsympathy only when the action is not understandable. And *vice versa.* That is, for Smith, for an action to be non-understandable it only means it is not optimal, i.e., the agent simply over-reacted or surrendered to temptation. So Smith, as stated earlier, conflates propriety with empathy, and impropriety with no-empathy. So, for him, to be unable to empathize arises only when the action is non-optimal. It is not because of vengeance, malice, or hate. Therefore, like many modern thinkers, Smith's conceptual tools do not have the ability to comprehend or identify the problem of malice or evil.

6.2 Compassion

Compassion is defined as a particular fellow-feeling that may or may not arise when the intentionality axis is engaged but, unlike sympathy, when the rationality axis is suspended. The spectator in quadrant II, definitely expresses compassion if he finds the action of the principal to be understandable, i.e., to be motivated by wellbeing—while withholding judgment as to whether it also efficient (propriety) or suboptimal (impropriety). As mentioned above, if the action is suboptimal, then the spectator or the judge within would be unsympathetic. But insofar as we are in quadrant II, the rationality axis is suspended and, hence, judgment of rationality is not considered. This definition of compassion coincides with Martha Nussbaum's, where she emphasizes that the issue is not about blame (judgment of rationality) but about the reduction of suffering, i.e., improvement of wellbeing:

> The emotion of compassion involves the thought that another creature is suffering significantly, and is not (or not mostly) to blame for that suffering. It does not involve the thought that someone is to blame for that suffering. One may have compassion for the victim of a crime, but one may also have compassion for

> someone who is dying from disease (in a situation where that vulnerability to disease is nobody's fault). "Humanity" I take to be a similar idea. So compassion omits the essential element of blame for wrongdoing [Nussbaum in Sunstein & Nussbaum, 2004, p. 301].

On the other hand, the spectator may express revulsion or disgust if he cannot understand or empathize with the action/emotion of the other, even if such emotion involves suffering. For instance, one may not empathize with the suffering of a serial killer, following the fact that one did not empathize with the killer's intention. The revulsion arises for the same reason when one sees someone drinking crankcase oil of his automobile or eating repulsive meant. In all these cases, such actions are revolting because they reduce wellbeing. The reduction of wellbeing can be the wellbeing of others, as in the case of malevolence, or the wellbeing of the self, as in the case cults. If the action is revolting, the spectator would feel "rejection" towards the action—while again suspending the rationality axis.

In light of the TAE hypothesis, we can easily now distinguish malevolence from selfishness—two phenomenon that are commonly confused in the biological and economic literature on altruism [see Khalil, 2004]. Selfishness, as mentioned above, is an act that the spectator can understand because the intention is to enhance the wellbeing of current self, but when the optimal choice is to take care more of the interest of future self or of the interest of important other. As such, the spectator, or judge within, expresses unsympathy towards selfish actions—while still empathetic with them. This different from expressing rejection—when the spectator is unempathetic with acts of malevolence whose main purpose is to reduce wellbeing without even promoting the selfish choices that promote the wellbeing of the current self.

6.3 Indulgence

Indulgence is defined as fellow-feeling when both axes, contrary to sympathy, are suspended. Here, the spectator enjoys the action of the principal *neither* because he or she understands the intention behind it *nor* because he or she finds the action proper. Rather, the spectator enjoys it because it reminds him of his own experiences in life or his own past pleasures.

This meaning of indulgence coincides with Theodor Lipps's [1960] concept of the aesthetic experience, which he incidentally calls "empathy" [see Gladstein, 1987]. Lipps conceived the aesthetic experience as the projection of one's self-centered emotion on the viewed object:

> Esthetic enjoyment is a feeling of pleasure of joy in each individual case colored in some specific way and ever different in each new esthetic object—a pleasure caused by viewing the object [Lipps, 1960, p. 374].

So, the aesthetic experience is no different from infectious laughter where one laughs because one hears the laughter of others, i.e., as a result of the "imitation function" of fellow-feeling, without really *understanding* the cause of their laughter.

Other examples of indulgence include the spoiling of a child with some pleasures because it allows the parent to enjoy, vicariously, the pleasures of the child. So, the parent would not assess the propriety of the child's enjoyment or the intentionality of the child's action. The parent (spectator) would provide resources to the child (principal) mainly to maximize the parent vicarious utility function.

Indulgence is not limited to enjoyment. Indulgence can involve pain, such as squeamishness, which the agent tries to avoid. For instance, one can be squeamish, and not eat meat, after a visit to the slaughterhouse. One would not eat meat for a week or a month not because

one is repulsed, but rather because the thought of blood reminds one of unpleasant experiences.

Indulgence can involve pain which the agent, amazingly, seeks. The spectator may seek to learn about the suffering of others, not out of compassion (which implies empathy), but rather out of self-centered need to indulge in one's own suffering. As the earlier quote from Nietzsche attests, Christianity to him is the key to indulgence (which he calls "suffering") that saps one's ambition and one's will to excel.

As mentioned earlier, theorists as far apart as Hobbes and Becker have relied on indulgence utility to explain altruism. It is proper to call such theories of altruism egocentric. Smith criticized Hobbes's egocentric theory of altruism—a critique that equally applies to Becker's [Khalil, 2001]. Smith argued that sympathy—which also include its corresponding action, altruism—does not stem from ego-centric, "warm glow" pleasures. If it stems from egocentric fellow-feeling, how come, Smith asks, men can have fellow-feeling towards women in labor, when in fact they could never in their own person undergo such an experience. The fact that the man empathizes with the woman as a woman is because the man does not dwell in his self-centered station, but rather he transports himself to the station of the woman:

> A man may sympathize with a woman in child–bed; though it is impossible that he should conceive himself as suffering her pains in his own proper person and character [Smith, 1976, p. 312].

Smith is correct that altruism, once narrowly defined, cannot be based on indulgence. But this does not rule out that in other schemes of income sharing, the motive of the spectator, who shares income with the principal, might be indulgence, i.e., vicarious pleasure rather than altruism. And such indulgence, given the suspension of the intentionality axis, is usually

facilitated by resemblance of traits [Khalil, 2002b].[12] Obviously such resemblance of traits, as Smith notes, does not exist between the man and the woman in labor. But resemblance of traits, contrary to Smith, can be the basis of schemes of income sharing other than altruism. Such other schemes include the spoiling of a principal—such as a child—in order for the spectator to indulge himself or herself.

6.4 Adulation

Adulation is defined as the fellow-feeling that may or may not arise when the intentionality axis is disengaged while, unlike indulgence, the rationality axis is engaged. Here, the spectator is enjoying, or suffering with the principal—not because, again, the spectator empathizes/rejects the intention of the principal. The issue of intentionality is irrelevant. Rather, the spectator expresses fellow-feeling because, similar to indulgence, he relates to what is happening to the principal as happening to his own personal station. The spectator is not examining the cause or incentive behind the principal's action. In this context, the spectator does not care about the principal *in his station*. The spectator is just observing the action of the principal in order to use it vicariously, i.e., as a vehicle for him to

[12] Actually, the resemblance of traits plays an important role in Hume's theory of sympathy. As David Levy and Sandra Peart [2004] show, Hume defines sympathy as indulgence and, in turn, argues that such indulgence is mediated through the resemblance of traits. However, Levy and Peart proceed and advance an interesting, although an indefensible thesis: Hume's notion of sympathy, i.e., grounded on resemblance of traits of one race vis-à-vis the rest, *necessarily entail* a narrower sense of civil society than Smith's notion of sympathy, i.e., grounded on humanity. It is correct that Hume advances a narrower notion of civil society than Smith. Further, it is correct that Hume advances the thesis as reconstructed by Levy and Peart. However, as Appendix A maintains, the thesis that the boundary of one's civil society is a *function* of sympathy even when understood as indulgence—whereas sympathy or indulgence implies justice (i.e., the reciprocal respect of property rights)—is simply indefensible.

boost his own self-evaluation..

There is one important difference between adulation of quadrant IV and indulgence of quadrant III. Indulgence is about the contagion of pleasure or gloom. Pleasure escalates as a result of being around people who are happy; gloom escalates as a result of being around people who are gloomy. Hume called the escalation "disagreeable sympathy" ("disagreeable" because Hume only examined the escalation of gloom). In contrast, adulation involves judgment of propriety that actually focuses on the spectator as a result of suspending the intentionality axis, . Such judgment takes the achievement of others not as they saw them, or whether they are satisfied with them. Rather they examine the achievement of others in order to compare them to their own actions and projects.

So, the spectator is not interested in how satisfied is the principal, i.e., how the action is related to the intention. The spectator is rather self-centered. The spectator is evaluating his or her own standing in relation to the standing of coveted positions of others. Such positions are the externalized reference point for what one believes to be his ability and his desire. Such believes are about the self and called elsewhere "noncognitive beliefs" to distinguish them from "cognitive beliefs" concerning the comprehension of the environment [Khalil, 2007d].

If one's quest or desire concerns etiquettes, the spectator asks whether the way the principal walks, eats, dresses, and so on, is more elegant than the way he or she walks, eats, dresses, and so on. The evaluation of one's etiquettes is not trivial as supposed at first look. It indicates one's care about health, risk, and so on. One's quest or desire can be wealth, knowledge, beauty, sociability, and so on. Whatever is the metric, the spectator measures his accomplishments in relation to the

principal's or, what is the same thing, in relation to his own goal. Both yardsticks are the same. The spectator, after all, selects the principal, or the social reference group, against which he or she would like to gauge his own performance.

If the spectator finds that the principal, with regard to the selected metric, has a higher achievement as a result of prudence and tenacious effort, the spectator would use the principal as an exemplar in order to exercise similar prudence and tenacity and achieve a similar standing. Such judgment of standing or status is more involved. Factors such as luck and natural aptitudes play a role [Khalil, 1996], which we will ignore here for simplicity. We will focus only on how the spectator judges relative standing, as if accomplishment is purely the outcome of prudence and tenacity. Our spectator may experience jealousy towards the principal as the spectator tries hard to attain his own desire that he sees so perfectly achieved by the object of his jealousy, the principal. The jealousy, though, is usually mixed with adulation especially when the jealous spectator starts to believe that he cannot attain what he truly desires.

On the other hand, if the spectator finds that the principal has a lower standing, i.e., the principal failed to act prudently and tenaciously, the spectator would feel pre-eminent or has a lead over the principal. Such feeling is called here "authority" in the sense that the spectator regards himself as superior vis-à-vis the principal with regard to the selected metric. The term "authority" is not used here in a pejorative sense as patronizing or condescending attitude on the part of the higher standing agent. Such patronizing attitude can develop, called below "pity." While pity presupposes authority, it involves another element, viz., elitism, as discussed below. The term "authority" is rather used here in the sense of mentoring, parenting, or acting as exemplar of propriety. Such

mentoring may involve the development of the principal's utility function. So the mentoring is not simply about increasing efficiency in the sense of providing either information or precommitments to assist the principal with self-command (prudence) in the face of temptations. So authority or mentoring differs from Cass Sunstein and Richard Thaler's [2003] notion of "libertarian paternalism" insofar as libertarian paternalism is simply about enhancing efficiency by providing either better information or precommitments.

So, the spectator in quadrant IV can express, towards the principal, the fellow-feeling of jealousy that may lead to adulation (in case the principal acts with propriety and tenacity) or authority (in case the principal is lacking in achievement). The adulation/authority twin perform same function: the assessment of one's self or others as one pursues the object of desire.

René Girard [1972] regarded, reminiscent of Nietzsche, *desire* as the defining question of the human condition. Broadly speaking, except for the theories of Nietzsche and a few others, modern social theory has neglected the role of desire as the entry point of theorizing about the human condition. Of course, most theories eventually discuss desire. But the point is whether desire acts as the organizing principal to make sense of diverse phenomena. Modern social theory is mainly concerned with the social contract in light of competing interests and the problem of free-riding. But humans still have to deal with desire even if they live as Robinson Crusoe. Girard's work show how the frustration of desire leads mortals to make Gods of each other. And in this act of adulation/authority, the lower status agent do not want to know that the emperor or the Gods have no clothes, as much as the people acting as authority do not want reveal themselves naked.

Karl Marx [1973] discussed at length adulation/authority relation that binds the chieftains,

kings and emperors with their subjects. Marx, though, restricted such adulation to pre-capitalist social formation. Marx argued that the root of such adulation, which I called "rank fetishism," is the fear of nature [se Khalil, 1992]. Marx was typical of modern social theory. He thought that adulation would whither away with the rise of capitalist mode of production because of technological progress, what he called the advancement of "forces of production." The advancement allows humans to control nature. Consequently, humans would no longer be scared with the rise of forces of production (technology). Thus, they would no longer tend to make Gods of mortals.

Smith was not as a modernist as Marx. He did not think that concern with rank and status would vanish with the rise commercial society. In his analysis of the origin of rank, Smith [1976, pp. 50-62] rather anticipates Nietzsche and Girard. This is not the place to elaborate his theory of authority, which challenges directly social contract theory of the socialist tradition of Jean-Jacques Rousseau as well as of classical liberalism stretching from Thomas Hobbes, John Lock, to James Buchanan [see Khalil, 1998; 2002a, 2005]. Stated briefly, for Smith, humans, *all* humans, would choose death over living lives that is empty of desire, i.e., the ambition to reach higher ranking goals. But most humans realize that they cannot attain their desire. So, they adulate other humans that seem to them more successful than them. Such adulation, the lower-rank spectators are not really sympathetic with the welfare of the rich and famous—because they are not engaging the intentionality axis. The spectators are rather operating from their own, self-centered fellow-feeling. So, the news about the more successful agents, i.e., the ones judged to embody desired goals according to the rationality axis, become the object of vicarious enjoyment.

Smith seems to be aware that such vicarious enjoyment, i.e., adulation, is different from his concept of sympathy. This is the case because he called adulation "imaginative sympathy." But Smith never tried to connect his concept of sympathy with adulation. In an earlier analysis, I called adulation in Smith "vicarious sympathy" [Khalil, 2002a, 2005]. I thought the term "vicarious sympathy" is better indicative of adulation than Smith's "imaginative sympathy. However, the term "vicarious sympathy" is, in light of the TEA hypothesis, is inadequate if we want to distinguish "adulation" from "indulgence"—since indulgence (quadrant III) also involves vicarious enjoyment.

To illustrate the difference between adulation and sympathy, let us examine the enormous "sympathy" accompanying the imprisonment of Paris Hilton. Ms. Hilton, a 26-year old heiress of the Hilton hotel fortune, is famous for being famous. So her achievement is not actually examined on her own station; they are rather the fancy of spectators of quadrant IV. The picture of Ms. Hilton splashed the front pages of newspapers around the world as she arrived, in early June 2007, at the Century Regional Detention Facility in suburban Los Angeles to serve a 45-day prison sentence for violating probation in an alcohol-related reckless driving case. Why all this interest and commotion for a 26-year old woman going to jail, in which she is expected to serve 23 days? Why even a website was set up on her behalf by fans to start a petition asking the Governor of California to pardon her? Is it sympathy based on the intentionality axis? As painful as the jail ordeal would be for her, there are more horrific ordeals that women undergo everyday in Southeast Asia with the slave-sex trade, and the more agonizing ordeals that women undergo in many poor African countries, viz., they have to take care of family members who have AIDS while they themselves are also infected with AIDS. If the fans of Ms. Hilton, and the wider public, are motivated by sympathy,

they would have instead spent their resources on the problems of sex-slave trade and AIDS.

Likewise, adult men cry when Princess Diane died—when they did not cry or did not feel the same intensity of loss when they lost their parents. The fascination with celebrity cannot stem from sympathy, as Smith long ago noted. It must be related to frustrated desire, where there is a judgment of what one can desire. Such a judgment leads to the ranking of people, where the higher rank is worshiped and venerated. Marx was wrong. The advancement of capitalist production failed to free us from rank fetishism and status inequality as Marx predicted. (In contrast, Marx predicted that the advancement of socialism would free us from *income* inequality).

Status inequality arouses the lower rank person to adulate the higher rank. The higher rank, if generous, usually reciprocates with authority as discussed above. Authority need not entail condescension, patronizing behavior, arrogance, or what is called below "pity" on the part of the higher ranked person.

But in many cases authority may lead to pity fellow-feeling. In this case, lower status people are not only expected to adulate the higher status ones, but they are also supposed to venerate them to the point of obsequiousness. When higher status agents express pity, lower status agent should express what is called here "obsequiousness." The twin fellow-feeling of *pity/obsequiousness* is not the product of simple status inequality. Simple status inequality generates the twin fellow-feeling of *authority/adulation*. It is conjectured here that the pity/obsequiousness twin is rather the product of status inequality mixed with another institution, called here "elitism." Elitism can be defined as the institution or ideology that people of lower rank are condemned to stay at a lower rank. These lower rank individuals are supposed to be almost confined to their station in life despite any effort they

spend.

Once status inequality is mixed with elitism, which need not be the case, the compound is usually the caste system, racial segregation, or stiff social segregation based on other kind of group identity such as religion, ethnicity, accent, and so on. As an example, the racial segregation setting of the novel *To Kill a Mockingbird*, mentioned at the outset, is such a compound of status inequality and elitism. Such racial segregation is not new in history. In fact, most societies, such as the caste system of India, are characterized by an institutional matrix of status inequality and elitism. Once seen in this light, such stiff institution of social segregation differs from racial hatred which can spawn episodes such as the holocaust. The distinction cannot be demarcated without the TAE hypothesis. As elaborated below, the intentionality axis is engaged in the case of racial hatred, but not in the case of racial or other kind of segregation.

Tom, given the institutional matrix of racial segregation, should not only adulate white people. He should also bow to them obsequiously. In return, white people would express not only authority, but also pity. So, pity can be defined as the fellow-feeling of authority that is mixed with patronization or elitism. The pity fellow feeling implies that the principal is not only of lower status but also he is cannot, too bad for him, rise higher in status because of his race, primitive beliefs, or backward culture.

While Smith used the term "pity," he used it interchangeably with "compassion" to denote general fellow-feeling. Nonetheless, Smith did not miss an opportunity to describe and criticize ostentatious and arrogant behavior, which is responsible for pity, and its twin fellow-feeling, obsequiousness. To wit, the terms "ostentatious" and "obsequious" are often encountered in *The*

Theory of Moral Sentiments. To note, Smith did not delineate between "authority" and its mutilated form, "pity" as conjectured here—i.e., delineate between status inequality, on one hand, and status inequality mixed with arrogance or elitism, which is responsible for social segregation, on the other.

For Smith [1976, pp. 255-259; see Khalil, 1996], arrogance is found in people who are inflicted with "weakness of character," i.e., people who are anxious about their standing in the pecking order of society. Smith skillfully distinguished between two "flavors" of arrogant, weak men: the "vain man" and the "proud man." Both exhibit self-aggrandizement. While the vain man is too ready to display his accomplishments in order to remind lower-ranking agents that they cannot reach his rank; the proud man is even too proud even to bother to display his accomplishments.

Status inequality, in short, has to be mixed with elitism to produce self-aggrandizement of the vanity type or the proud type [see Khalil, 2000a]. Status inequality, which engenders the adulation/authority twin, can take place without elitism. Status inequality need not be thrown away along with elitism in order to undermine the obsequiousness/pity twin. To throw away status inequality with elitism would amount to throwing the baby away with the bathwater. Humans may never be able to avoid status inequality. The commercial society (or any vibrant society) might be free of elitism and the caste system. This does not mean, as Marx assumed, that the commercial society is also free of status inequality. In any case, we need the concept of status inequality and, correspondingly, the adulation/authority fellow-feeling if we ever want to make sense of *assabiya*.

7. How far Apart is *Assabiya* from other "Social Preferences"?

The proposed TEA hypothesis allows us to identify adulation, the fellow-feeling of quadrant IV, as different from sympathy, the fellow-feeling of quadrant I.

Another payoff of the TEA hypothesis is now almost obvious. When the spectator experiences adulation, it is complemented usually by the principal expressing authority. The adulation/authority complementary emotions might be, it is conjectured here, the elementary building block for the study of allegiance, group solidarity, or the bond that unites the citizens of the state. In fact, Adam Smith [1976. pp. 50-62] argued that status inequality, i.e., the adulation/authority twin, is the corner stone of understanding political authority.

Smith [1978] spent a great deal of effort in analyzing the nature of political authority. Smith directly criticized John Locke's theory [Khalil, 1998]. The social contract idea, based on interests, simply misses the role of desire and, hence, fails to grasp the nature of authority. For Smith [1976, p. 50] desire is the entry point of analysis if we want to explain authority. What matters for Smith's analysis is that desire is often frustrated. Frustrated desire is nonetheless fulfilled though adulation, as discussed above. Adulation amounts to the fusion of egos, where the spectator identifies his ego with the imagined ego of the team, producing what is usually called "team spirit." But team spirit may not be different from how a sports fan identifies with a sports team or a movie viewer identifies with the hero of a film. So, the fusion of egos cannot be the whole story of political allegiance. To wit, Smith argues that there is another element, aside from authority, that is needed in order to explain political allegiance.

Stated briefly, the adulation/authority twin must be combined with the principle of interest or utility [see Khalil, 2002a, 2005]. Once authority is combined with interest, the adulation/authority

twin is transformed into allegiance, group solidarity, or, in short, a*ssabiya*. Such *assabiya* prompts spectators to cry when they see their king, touch their flag, or hear the national anthem.[13]

How can we explain such nationalist or *assabiya* emotion? Is it the same as altruism and fairness, which also benefit other group members? Actually, motives such as *assabiya* (under the guise of group identity), altruism, and fairness have been receiving great attention in the literature under terms such as "social preferences" and "prosocial preferences" [e.g., Gintis, 2003; Bowles, 2004; Gintis *et al.*, 2005; Bénabou & Tirole, 2006]. One has to be careful, though, not to suggest that altruism and fairness have the same standing as *assabiya*, and just lump them all as "social preferences." In light of the TAE hypothesis, we should not use the same model to conceive *assabiya* as the once used to conceive altruism and fairness.

Stated briefly, altruism and fairness are ultimately about the evaluation of action in light of the intentionality axis. That is, the agent is trying, in both altruism and fairness, to enhance wellbeing as evaluated by the intentionality axis. In contrast, with the adulation/authority twin, responsible for *assabiya*, the intention of the actor is not under consideration to start with. It does not matter whether the principal's action is about wellbeing or it is not. Rather, what matters is how a lower status spectator feels by imagining the accomplishments of the great and powerful as it they

---

[13] In fact, one major aspect of Smith's notion of the invisible hand in *The Theory of Moral Sentiments* [Khalil, 2000b] is about the spontaneous rise of political order. He discusses a length how the myopic sentiments of adulation/authority gives rise, once combined with interest, to allegiance. Such allegiance affords political order which would whither away if there is no *assabiya*.

are his or her own. So, when a spectator acts according to allegiance, the action is not the same as when he or she acts altruistically or fairly.

8. Towards a Theory of Evil

The opposite of altruism is selfishness. The literature in economics, as well as socio-biology [e.g., Wilson, 1975], uses terms such as "spite" and "malevolence" as also the opposite of altruism. This implies that "selfishness" and "malevolence" are interchangeable emotion/action.

This should not be surprising. We can only distinguish selfishness from malevolence if we have a theory of evil. Such a theory is not possible if we fail to separate the intentionality axis from the rationality axis.

It is no exaggeration that "evil" as a concept was almost eradicated from modern social science theory with the rise of social contract theory and especially with the advance of Jeremy Bentham's utilitarianism and his cost-benefit view of crime and punishment. However, there are many exceptions. One exception, already mentioned, is Girard's theory of desire. While desire can be the seat of creativity and entrepreneurship, a frustrated desire can engender the feeling of hopelessness and lead to envy. But most people chose some intermediate existence of identifying themselves with the successful, glamorous, famous, and rich. In his study of novels, Girard shows how agents make God of each others, venerate and adulate heroes, and so on, as approximations of their deep-held desires. As discussed earlier, such adulation is the basic block for the establishment of authority and, with further conditions, political allegiance, at least for Adam Smith. But ones who, out of pride, would not create idols and submit to Gods, but still feel hopeless in improving

their conditions, would usually slip into envy, hate, and maybe evil.

It can be the case that the agent does not want to approximate his desire by hero worship or adulation. If this is combined with a state of hopelessness, where the agent does not believe he can satisfy his desire on his own, it leads to frustrated desire. A frustrated desire is the root of envy.

In the tradition of Girard, Dupuy [2004, 2006], in his analysis of Adam Smith's moral sentiments, grouped jealousy with envy and called them "invidious" sentiments. However, it is important to distinguish jealousy from envy. In jealousy, the principal does not participate in the adulation of superiors because he thinks that they are his equals, and he works hard to prove it. Even when he gives up, and embarks on adulation, the agent is in a sense still full of hope. Further, even when the consequent status inequality is mixed with elitism, and the product is social segregation, the system is still based on hope.

This is not the case with envy. Here the agent gives up hope. He does not think that he is capable of improving his standing. So, he resorts to another kind of pleasure, the pleasure of seeing others getting hurt, i.e., malice and envy. Such sentiments are invidious, whereas jealousy is not. Jealousy invites hard work and entrepreneurship, or to vicariously enjoy it through admired others.

To understand evil, and to differentiate racial hatred from social segregation, we need the TAE hypothesis, i.e., we need to engage the intentionality axis as separate from the rationality axis. While an act can be judged as rational, the question of intention remains. The spectator judges an intention as evil when it evokes revulsion rather than empathy. For revulsion to arise, the spectator must have become cognizant that the principal intends to reduce wellbeing as a way to promote his own sense of accomplishment.

But how can an evil act, i.e., an act that sabotages wellbeing, differ from selfishness? Two clarifications are needed. First, the principal can reduce total welfare without resorting necessarily to sabotage, i.e., lowering the budget constraint of the envied subject. The principal can choose an obsessive belief that engenders and nurtures hatred of others, as the case when he joins a cult. Such an obsessive belief would effectively make him unable to utility his own budget, rendering most of his resources destroyed or frozen. So, cults can be defined as self-directed acts of evil insofar as they arrest development. However, to keep the analysis simple, we will assume that an evil act is usually vengeance that reduces the budge of envied subject, not the self.

Second, it is possible that the act of vengeance may involve some wealth transfer from the object of envy to the principal. But the forced transfer, even when significant, is not the main motive behind sabotage. The main motive of sabotage is not transfer of wealth—which mostly if not all rendered useless through destruction. This is the case because the main motive behind vengeance or sabotage is exactly the destruction of what the envied subject, which allows the principal to bolster his utility via the framing effect.

Such an act of evil cannot be understood unless the intentionality axis is separated from the efficiency axis. The act of evil can be efficient—when the principal's manipulation of the framing effect delivers the desired outcome. It can also be suboptimal—when the principal's manipulation is botched due to impatience and myopia. So, to model evil, we need to ignore the efficiency axis by assuming that the principal would act optimally.

It would be outside the scope of this paper to build a model of evil. It is sufficient to state that we need to distinguish two kinds of ego-utility. One ego-utility arises from assessing one's own

action, what was called earlier "integrity" or, in general, "self-satisfaction utility." The other, called "malevolent utility," arises from assessing one's own wellbeing by sabotaging what one could become. This may entail the reduction of wellbeing of envied other or, in the case of cults, the reduction of wellbeing of the self. But we cannot distinguish between these two kinds of ego-utility if the intentionality axis is conflated with the rationality axis.

9. Conclusion

This paper's central contribution is the "two-axis evaluation" hypothesis. Namely, the rationality axis and the intentionality axis are orthogonal. The TAE hypothesis is proposed to solve the mirror-neuron paradox.

As expressed by David Hume long-ago, there is a contradiction between the "understanding function" of mirroring and the "imitation function" of mirroring. The "understanding function," what Hume calls "agreeable" sympathy, leads to the attenuation of the principal's emotion/action. The "imitation function," what Hume calls "disagreeable" sympathy, leads to the escalation of the principal's emotion/action. When he learned that Smith is preparing a second edition of *The Theory of Moral Sentiments*, Hume challenged Smith: How could the two functions, which are contradictory, arise from the same primitive emotion, viz., mirroring?

To solve this mirror-neuron paradox, this paper conjectures two orthogonal institutions or evaluations: rationality axis versus intentionality axis. The rationality axis asks: Is the action rational or is it suboptimal? The answer allows us to judge whether the action is proper or improper. In contrast, the intentionality axis asks: Is the intention is wellbeing or is it malevolence (evil)?

The answer allows us to judge whether the action is subject to understanding (empathy) or revolting (disgusting).

Matters are actually more complex. Mirroring can be processed while either axis, or both, is frozen or suspended. That is, mirroring can take place by asking one question while the other is suspended—or by suspending both questions. This gives rise to four possible faces of fellow-feeling: sympathy, indulgence, compassion, and adulation. When rationality axis and intentionality axis are invoked, we might have sympathy or unsympathy; when neither is invoked, we have indulgence; when one or the other axis is invoked, we might have either compassion or indulgence. In this light, sympathy attenuates the principal's emotion because, to be judged approvingly, the principal has to lower the pitch of his emotions. This is not the case with indulgence, where there is no judgment to start with, which leads to the reinforcement of the principal's emotion.

Economists have exclusively focused on the rationality axis—totally ignoring the intentionality axis. There is little hope that economists will tackle the intentionality axis in the near future. To start with, economists take the preferences of agents as given. It is not up to a scientific program to assess the intentionality of agents, not to mention the analysis of evil.

Even with the rise of behavioral economics, there is little hope to tackle the question of evil. Behavioral economists are challenging, among other things, the revealed preference axiom. But even if one disputes the axiom, and admits that agents do not behave rationally, this does not invoke the intentionality axis. The challenge of behavioral economics is rather restricted to the rationality axis.

The major result of this paper is that the quest after the intentionality axis cannot be reduced to the rationality axis. Such a reduction has given rise to the mirror-neuron paradox, i.e., make us unable to distinguish understanding from imitation. Such inability hinders us from theoretically distinguishing sympathy, from compassion, indulgence, and adulation. The distinction among these four kinds of fellow-feeling is essential for modeling altruism, fairness, and group solidarity (*assabiya*). The solution of the paradox also allows us to distinguish selfishness from malevolence.

Economics is not alone in ignoring the intentionality axis. It is actually the mark of the rise of modern social science—on the shoulders of Machiavelli, Hobbes, Locke, Rousseau, and Marx—to be antagonistic to the question of evil. Modern social science views the "economic problem" or the "human condition" as about the engineering of the best institutions that concern issues raised exclusively by the rationality axis, i.e., the benefits of competition and cooperation among rival or complementary interests. So, if one person hurts another it is only because the perpetrator is pursuing, efficiently or suboptimally, his or her wellbeing. So, we have atoms that collide simply as a result of the pursuit of wellbeing. The modern palace of social science has generally no room for the question of evil. It is hoped that this essay has opened a window in the palace that is wide enough to entice further scientific study of evil and other aspects of the intentionality axis.

Appendix A:
Sympathy and Enslavement

David Hume focused greatly on the resemblance of traits as the basis of sympathy, as David Levy and Sandra Peart [2004] show. As Hume puts it:

> Now 'tis obvious, that nature has preserv'd a great resemblance among all human creatures, and that we never remark any passion or principle in others, of which, in some degree or other, we may not find a parallel in ourselves. The case is the same with the fabric of the mind, as with that of the body. However the parts may differ in shape or size, their structure and composition are in general the same. There is a very remarkable resemblance, which preserves itself amidst all their variety; and this resemblance must very much contribute to make us enter into the sentiments of others, and embrace them with facility and pleasure. Accordingly we find, that where, beside the general resemblance of our natures, there is any peculiar similarity in our manners, or character, or country, or language, it facilitates the sympathy. The stronger the relation is betwixt ourselves and any object, the more easily does the imagination make the transition, and convey to the related idea the vivacity of conception, with which we always form the idea of our own person [Hume, 1896, p. 211].

It seems that once one adopts the definition of sympathy as about indulgence, as Hume does, one would advance the postulate that sympathy arises only if the spectator and the principal share a set of common traits [see Khalil, 2002b].

Levy and Peart proceed and advance an interesting, although indefensible thesis: Hume's notion of sympathy, i.e., sympathy is grounded on resemblance of traits of one race vis-à-vis the rest, *necessarily entail* a narrower sense of civil society than Smith's notion of sympathy, i.e., sympathy is grounded on humanity.

To analyze this thesis, let us first define the term "civil society" (CS). Every CS has a boundary, where rules of justice are valid to the members of the CS under focus while they are invalid with regard to outsiders. So, while members of a society cannot breach the property

rights of other members of the same society, they are justified in exploiting and enslaving members that belong to other civil societies. This is expressed in tribal religions, where each tribe or nation has its particular God or Gods, while recognizing that outsiders have their own. When one regards the whole humanity as belonging to a single civil society, no one state can enslave another state.

But still one can justify the enslavement of nonhuman animals on the basis that they fall outside one's civil society, as John Rawls's [1971] theory of justice implies [see Sunstein & Nussbaum, 2004]. For Rawls, the concept of "animal rights" has no meaning because CS cannot be broader than humankind. Nonhuman animals cannot enter into a contract with humans. So, for Rawls, it is justified to exploit and enslave nonhuman organisms. Along the same line of argument, if one's civil society is limited Europe, the Europeans are then justified in exploiting and enslaving non-Europeans, and *vice versa* [see Khalil, 2007b].

While Hume's civil society is limited to one's own countrymen, where people recognize common resemblance, Smith's civil society encompasses the whole humanity. So, Hume justifies the European enslavement of the natives of the Americas on the basis that the native Americans are outsiders of the European civil society. In contrast, Smith defines civil society to include all humanity.

The question is the following: Can the difference between Smith and Hume concerning the boundary of civil society be traced, as Levy and Peart argue, to their different views on fellow-feeling. While Smith views fellow-feeling as sympathy, while Hume views it as indulgence. The question is valid even if Hume, as well, thinks that boundary of civil society is

a function of indulgence (as defined here), i.e.,

$$CS = CS[I(t)]$$

where I is indulgence, and t common traits. According to this theory of CS, the institution of property rights, which defines CS, is ultimately based on the set of common traits such as common culture, religion, values, and so on. This theory is not uncommon. It is, in fact, the driving force behind much of culturalist economics, as exemplified by Douglass North [2005; Khalil, 2007c].

As the literature on the MNS shows [Buccino *et al.*, 2004; Rizzolatti & Craighero, 2004], sympathy takes place among nonconspecifics, i.e., organisms that even do not belong to the same species. A dog understands the action taken by a monkey, such as reaching for an object, and *vice versa*. But even if we define sympathy-as-indulgence, as Hume does, and hence involves the "imitation function" rather than the "understanding function," a human can still imitate other humans that fall outside his group. An ego-centered person, involved in indulgence, can even become stimulated by the emotions of nonhuman organisms if he or shed happened to construct a *broader* set of resemblance of traits. For instance, one can find mammalian resemblance set that lumps humans with all other mammals. Or one can find vertebra resemblance set that lumps humans with all nonhuman vertebra such as fish. Following Hume's CS theory, we may include all mammals within the human civil society by simply expanding the set of common traits to be mammalian. To wit, to push the argument further, if one takes the common traits to be vertebrae, then only non-vertebrae organisms fall outside the civil society that includes humans.

So, the determination of boundary of society cannot, logically, be derived from the issue of resemblance of traits. The set of resemblance of traits is an elastic concept. Thus, the fact that Hume has a narrower view of civil society cannot be grounded on his view of sympathy—even if he thought so. His view of sympathy allows CS to be as broad as to encompass all animals, including insects, and to be as narrow as to encompass only the people who live in his neighborhood.

In fact, in response to Rawls, Peter Singer [2002; in Sunstein & Nussbaum, 2004] employs Hume's logic. Namely, Singer advocates the rights of many animals on the basis of fellow-feeling, where such feeling can be stretched out beyond humankind. But again, there is no hard boundary for CS as much as there is no hard boundary to the limits of fellow-feeling. It might be just the case that the issue of CS is unrelated to the issue of fellow-feeling. If so, how far we want to extend protection to animals it is up to how far we want to extend our fellow-feeling [see Posner, in Sunstein & Nussbaum, 2004].

References

Arrow, Kenneth J. "Extended Sympathy and the Possibility of Social Choice." *American Economic Review*, Papers and Proceedings, February 1977, 67:1, pp. 219-225.

Basch, Michael. "Empathic Understanding: A Review of the Concept and Some Theoretical Considerations." *Journal of the American Psychoanalytic Association,* 1983, 31:1, pp. 101-26.

Becker, Gary S. "Altruism in the Family and Selfishness in the Market Place." *Economica*, February 1981, 48, pp. 1-15.

_____. "A Note on Restaurant Pricing and Other Examples of Social Influences on Price**."** *Journal of Political Economy*, October 1991, 99:5, pp. 1109-1116.

_____. *Accounting for Tastes*. Cambridge, MA: Harvard University Press, 1996.

_____ and Kevin M. Murphy. "A Simple Theory of Advertising as a Good or Bad." Quarterly Journal of Economics, November 1993, 108:4, pp. 941-964.

Bénabou, Roland and Jean Tirole. "Incentives and Prosocial Behavior." *American Economic Review*, December 2006, 96:5, pp. 1652-1678.

Bernheim, B. Douglas and Oded Stark. "Altruism Within the Family Reconsidered: Do Nice Guys Finish Last?" *American Economic Review*, 1988, 78: 1034–1045.

Binmore, Ken. *Game Theory and the Social Contract. Volume 1: Playing Fair*. MIT Press, 1994.

_____. *Game Theory and the Social Contract. Volume 2: Just Playing*. MIT Press, 1998.

Bowles, Samuel. *Microeconomics: Behavior, Institutions and Evolution*. New York: Russell Sage; Princeton, N.J.: Princeton University Press, 2004.

Buccino, Giovanni, Fausta Lui , Nicola Canessa, Ilaria Patteri, Giovanna Lagravinese, Francesca Benuzzi, Carlo A. Porro, Giacomo Rizzolatti. "Neural Circuits Involved in the Recognition of Actions Performed by Nonconspecifics: An fMRI Study." *Journal of Cognitive Neuroscience*, 2004; 16, pp. 114-126.

Burt, Austin and Robert Trivers. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, MA: Harvard University Press, 2006.

Darwall, Stephen. "Empathy, Sympathy, Care." In Darwall's *Welfare and Rational Care*. Princeton, NJ: Princeton University Press, 2002, ch. 3.

_____. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.

Dupuy, Jean-Pierre. "Intersubjectivity and Embodiment." *Journal of Bioeconomics*, 2004, 6:3, pp. 275-294.

_____ "Invidious Sympathy in *The Theory of Moral Sentiment*." *Adam Smith Review*, 2006, 2, pp. 98-123.

Fontaine, Phillipe. "Recognition and Economic Behavior: Sympathy and Empathy in Historical Perspective." *Economics and Philosophy*, 1997, 13, pp. 261-280.

_____. "The Changing Place of Empathy in Welfare Economics." *History of Political Economy*, Fall 2001, 33:3, pp. 387-409.

Frank, Robert H. *Microeconomics and Behavior*, 6th ed. New York: McGraw Hill Irwin 2006.

Fudenberg, Drew and David K. Levine. "A Dual-Self Model of Impulse Control." *American Economic Review*, December 2006, 96:5, pp. 1449-1476.

Gallese, V., C. Keysers, and G. Rizzolatti. "A Unifying View of the Basis of Social Cognition." *Trends in Cognitive Sciences*, 2004, 8:9, pp. 396-403.

Gintis, Herbert. "Solving the Puzzle of Prosociality." *Rationality and Society*, 2003, 15, pp. 155-187.

_____, Samuel Bowles, Robert Boyd, and Ernst Fehr (eds.). *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life.* Cambridge, MA: MIT Press, 2005.

Girard, René. *Deceit, Desire and the Novel: Self and Other in Literary Structure*, translated by Yvonne Freccero. Baltimore: The Johns Hopkins University Press, 1976.

Gladstein, Gerald A. "The Historical Roots of Contemporary Empathy Research." *Journal of the History of the Behavioral Sciences*, 1984, 20:1, pp. 38-59.

_____ *et al. Empathy and Counseling: Explorations in Theory and Research*. New York: Springer-Verlag, 1987.

Gordon, Robert. "Sympathy, Simulation, and the Impartial Spectator." *Ethics*, Summer 1995. (Reprinted in Larry May, Marilyn Friedman, and Andy Clark *(eds.) Mind and Morals: Essays on Ethics and Cognitive Science* Cambridge, MA: MIT Press, 1996.)

Haakonssen, K. "Introduction." In Adam Smith's *The Theory of Moral Sentiments*. Cambridge: Cambridge University Press, 2002.

Haig, David. "Genetic Conflicts in Human Pregnancy." *Quarterly Review of Biology*, December 1993, 68:4, pp. 495-532.

_____. "On Intrapersonal Reciprocity." *Evolution and Human Behavior*, 2004, 24, pp. 418-425.

Harsanyi, John. "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility." *Journal of Political Economy*, 1955, 63, pp. 309–321.

_____. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations.* Cambridge: Cambridge University Press, 1977.

Heidegger, Martin. *Being and Time,* trans. John Macquarrie, Edward Robinson. New York: Harper, 1962.

Hollis, Martin. *Trust Within Reason.* Cambridge University Press, 1998.

Hume, David. *A Treatise of Human Nature*, 3 vols, ed. By L.A. Selby-Bigge. Oxford: Clarendon Press, [1740] 1896.

Hurley, Susan and Nick Chater (eds.). *Perspectives on Imitation: From Neuroscience to Social Science*, 2 vols. Cambridge, MA: MIT Press, 2005.

Ibn Khaldûn. *The Muqaddimah: An Introduction to History*, 3 vols., 2nd edition. Trans. From the Arabic by Franz Rosenthal, Bollingen Series XLIII. Princeton: Princeton University Press, 1967.

Karni, Edi, and David Schmeidler. "Fixed Preferences and Changing Tastes." *American Economic Review, Papers and Proceedings*, May 1990, 80:2, pp. 262-267.

Khalil, Elias L. "Beyond Self-Interest and Altruism: A Reconstruction of Adam Smith's Theory of Human Conduct." *Economics and Philosophy*, October 1990, 6:2, pp. 255-273.

_____. "Nature and Abstract Labor in Marx." *Social Concept*, June 1992, 6:2, pp. 91-117.

_____. "Respect, Admiration, Aggrandizement: Adam Smith as Economic Psychologist." *Journal of Economic Psychology*, September 1996, 17:5, pp. 555-577.

_____. "Is Justice the Primary Feature of the State? Adam Smith's Critique of Social Contract

Theory." *European Journal of Law and Economics*, November 1998, 6:3, pp. 215-230.

_____. "Symbolic Products: Prestige, Pride and Identity Goods." *Theory and Decision*, August 2000a, 49:1, pp. 53-77.

_____. "Beyond Natural Selection and Divine Intervention: The Lamarckian Implication of Adam Smith's Invisible Hand." *Journal of Evolutionary Economics*, 2000b, 10:4, pp. 373-393.

_____. "Adam Smith and Three Theories of Altruism." *Recherches Économiques de Louvain – Louvain Economic Review*, 2001, 67:4, pp. 421-435.

_____. "Is Adam Smith Liberal?" *Journal of Institutional and Theoretical Economics*, December 2002a, 158:4, pp. 664-694.

_____. "Similarity vs. Familiarity: When Empathy Becomes Egocentric." *Behavioral and Brain Sciences,* February 2002b, 25:1, p. 41.

_____. "What is Altruism?" *Journal of Economic Psychology*, February 2004, 25:1, pp. 97-123.

_____. "An Anatomy of Authority: Adam Smith as Political Theorist." *Cambridge Journal of Economics*, January 2005, 29:1, pp. 57-71.

_____. "Introduction: Smith the Hedgehog." *Adam Smith Review*, 2006, 2, pp. 3-20.

_____. "Rank Fetishism and Corruption: Why the Adulation of Vacuous Celebrities can Enhance Productivity." A working paper, 2007a.

_____. "The Moral Justification of Enslavement." A working paper, 2007b.

_____. "Roadblock of Culturalist Economics: Douglass North and the Retreat of Economics." A working paper, 2007c.

_____. "The Cognitivist Fallacy: The Heckman Puzzle." A working paper, 2007d.

Lee, Harper. *To Kill a Mockingbird*. London: Mandarin, (1960) 1989.

Levy, David M. *How the Dismal Science Got its Name: Classical Economics & the Ur-text of Racial Politics.* Ann Arbor: University of Michigan Press, 2001.

_____ and Sandra J. Peart. "Sympathy and Approbation in Hume and Smith: A Solution to the other Rational Species Problem." *Economics and Philosophy*, October 2004, 20:2, pp. 331-349.

Lewis, Michael and Jeannette M. Haviland-Jones (eds.). *Handbook of Emotions*. New York: Guilford Press, 2000, pp. 637-653.

Lipps, Theodor. "Empathy, Inner Imitation, and Sense-Feelings." In Melvin Rader (ed.) *A Modern Book of Esthetics: An Anthology*, 3rd ed. New York: Holt, Rinehart and Winston, (1935) 1960, pp. 374-382.

Marx, Karl. *Grundrisse*, translated with a foreword by Martin Nicolaus. New York: Vintage, 1973.

_____. *Capital*, vol. 1, intro. by Ernest Mandel. Harmondsworth: Penguin and London: New Left Books, 1976.

McCloskey, Deirdre N. *The Bourgeois Virtues: Ethics for an Age of Commerce*. Chicago: University of Chicago Press, 2006.

Miller, William Ian. *The Anatomy of Disgust*. Cambridge, MA: Harvard University Press, 1997.

Nietzsche, Friedrich. *The Anti-Christ: Curse on Christianity*. In *The Nietzsche Reader*, ed. By Keith Ansell Pearson and Duncan Large. Malden, MA: Blackwell, 2006.

North, Douglass C. *Understanding the Process of Economic Change*. Princeton, NJ: Princeton University Press, 2005.

Peart, Sandra J. and David M. Levy. "Sympathy and its Dicontents: `Greatest Happiness' versus the 'General Good'." *European Journal of History of Economic Thought*, Autumn 2004, 11:3, pp. 453-478.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

Rizzolatti, Giacomo, L. Fadiga, L. Fogassi, and V. Ballege. "Resonance Behaviours and Mirror Neurons." *Archives Italiennes De Biologie*, 1999, 137, pp. 88-99.

Rizzolatti, Giacomo and L. Craighero. "Mirror Neuron: A Neurological Approach to Empathy." A working paper, University of Parma, Parma, Italy, 2004a.

_____. "The Mirror-Neuron System." *Annual Review of Neuroscience*, 2004b, 27, pp. 169-192

Rozin, Paul, Jonathan Haidt, and Clark R. McCauley. "Disgust." In Michael Lewis and Jeannette M. Haviland-Jones (eds.) *Handbook of Emotions*. New York: Guilford Press, 2000, pp. 637-653.

Rustichini. Aldo. "Neuroeconomics: Present and future." *Games and Economic Behavior*, 2005, 52:2, pp. 201-212.

Scheler, Max. *The Nature of Sympathy*; trans. by Peter Heath, intro. by W. Stark. New Haven, CT: Yale University Press, 1954.

Sen, Amartya K. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy & Public Affairs*, 1977, 6:4, pp. 317-344.

Singer, Peter. *Animal Liberation*. New York: Ecco, 2002.

Smith, Adam. *The Theory of Moral Sentiments*, D.D. Raphael and A.L Macfie (eds), Oxford: Oxford University Press, 1976.

_____. *The Correspondence of Adam Smith*, ed. E.C. Mossner and I.S. Moss. Oxford: Oxford University Press, 1977.

_____. *Lectures on Jurisprudence*, edited by R.L Meek, D.D. Raphael, and P.G. Stein. Oxford: Clarendon Press, 1978.

Stein, Edith (Saint Teresa Benedicta of the Cross Discalced Carmelite). *The Science of the Cross*, trans. by Josephine Koeppel. Washington, DC: ICS Publications, 2002.

_____. *On the Problem of Empathy*, 2nd ed. translated by Waltraut Stein, foreword by Erwin W. Straus. The Hague: M. Nijhoff, (1917) 1970.

Sugden, Robert. "Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling." *Economics and Philosophy*, 2002, 18:1, pp. 63-87.

Sunstein, Cass R. and Richard Thaler. "Libertarian Paternalism is Not an Oxymoron." *University of Chicago Law Review*, Fall 2003, 70:4.

Sunstein, Cass R. and Martha C. Nussbaum (eds.). *Animal Rights: Current Debates and New Directions*. Oxford: Oxford University Press, 2004.

Thaler, Richard H. and Hersh M. Shefrin. "An Economic Theory of Self-Control." *Journal of Political Economy*, April 1981, 89:2, pp. 392-406.

Veblen, Thorstein. *The Theory of the Leisure Class*. New York: Modern Library, 1934.

Wilson, Edward O. *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press, 1975.