

Neurology and Economics in Moral Philosophy and the Motivation of Altruism through Fairness

Adham K. Mandour, Matthew Crippen, John Salevurakis and Edwin Cruz-Rivera*

* *Corresponding author*

Abstract

Despite the generality of philosophical, economic, and psychological arguments exploring human nature, resulting models based upon logical deduction have lead to disparate and contradictory theories of morality. In fact, an issue central to moral theory not yet resolved is whether altruism truly exists. Its mode of operation, doubts of its existence as a pure form of sacrifice, and whether selfish motivation exists as its true foundation have all been addressed in general and fundamentally abstract arguments both in favor and against altruism as an extension of human behavioral patterns.

However, empirical evidence from the biological sciences can provide more concrete evidence to explain altruistic behavior. Using scientific evidence available from functional magnetic resonance studies and economic game theory, we argue that altruistic behavior, as a basis of cooperation, is not motivated by self-serving gains. Altruistic behavior is ultimately a pre-requisite for cooperation, and punishment maintains it if necessary. We propose that, given the evidence provided here, competition between individuals (and the rational self-interest doctrine that drives it forward in modern society) is ultimately contrary to basic human motivation. To that end, this paper gives qualified credence to Marx's conception of an alternative social and economic reality denying the present perception of competition and conflict as virtuous.

Contrasting modern theories of morality

Philosophers broadly categorize consequentialist moral theories into ethical egoism, psychological egoism, and utilitarianism. Ethical and Psychological egoism stress that all moral actions are deemed good in terms of their consequence, the best of which is the one that serves an individual's self-interest. Within these frameworks, people are treated as means rather than ends-in-themselves, so long as their interests are congruent with one's own. Altruism for its own sake is rejected in ethical/psychological egoism, and also in utilitarianism, for both regard no action as good in for its own sake, but only good as evinced by the desirable consequences it produces. However, unlike ethical egoism, utilitarianism is not individual-centric *per se*, for it adopts the greatest common good for the greatest number of individuals as its primary objective. As such, altruism in its extreme form is permissible only if it brings greater happiness to the greater number of individuals thus offsetting any extant disutility that was generated by the action.. In the classical definition of utilitarianism as proposed by Jeremy Bentham and John Stuart Mill, and although Mill recognized "qualitative" elements of utility, the definition of moral goodness (particularly for Bentham) in that sense is strictly defined in terms of quantitatively measured happiness (Reviewed in Rachels 1986). Thomas Hobbes offered a contrasting view that aligns more with ethical egoism, arguing that self-interest is the main driving force behind all human action, including cooperation, thus ruling out the very existence of genuine altruism. In short, individuals in their natural state would sooner bludgeon one another for their own

individual gain was it not that their interests, either naturally or in a world governed by laws, existed in concert with one another. He explains cooperation under the scope of agreed upon law by stating,

Hereby it is manifest, that during the time men live Civil without a common Power to keep them all in awe, they are at condition which is called Warre; and such a warre; as is of everyman, against every man (Hobbes 1651/1909, p. 91).

Immanuel Kant proposed an entirely different origin and driver for morality of all previously stated theories. For Kant, rational beings are ends in themselves due to the freedom that rationality allows them to achieve, and morality ultimately manifests as a Universal Law arising from said rationality. In *The Groundwork of the Metaphysics of Morals*, he states, "Always treat humanity, whether in your own person or in the person of any other, never simply as a means but always at the same time, as an end" (Kant 1785/2005, p. 107).

The underpinning on this Universal Law is a series of Categorical Imperatives that are intrinsically valid as they arise from rationality and, thus, should be universally and unconditionally obeyed in accordance with the Law. Thus, actions are not measured by their ultimate effect or purpose, but rather acquire their value inasmuch as they conform to this Universal Law. The determination of the Imperatives is based on Kant's Formula of Universal Law, which requires the derivation of maxims of universal simultaneous applicability. It is therefore proper to, "Act as if the maxim of

your action were to become through your will a universal law of nature” (Kant 1785/2005, p.10).

Thus, morality can be interpreted as a natural extension of rational thought and not as an idiosyncratic manifestation of human efforts to maximize individual well-being: a deontological, strictly non-utilitarian perspective. This, of course, has strong bearing on altruism. By regarding others as ends in themselves, it can be concluded that altruism without derived benefit is a moral good, and for the sake of logical consistency with Kant’s form of universally applied norms, it is only morally pure if it brings no joy or pleasure to those who dispense it, otherwise it is simply an inclination and has no intrinsic moral worth. Thus altruism motivated only by empathy, or stemming from anticipated and symmetric reciprocity, is considered wholly amoral.

Dissecting Marxian Morality

Identifying the moral stance of Marx can be problematic. Marx was not a moral philosopher and did not align himself with any of the previously reviewed moral systems. Further, he himself claims that scientific socialism “abolishes all morality instead of constituting it on a new basis” (Marx, Engels 1906, p. 26). This presumption was based on the premise that the prevailing morality at any given time is a construct of the ruling class, and one that shifts with the rise and fall of ruling classes throughout history. Marx held the view that social superstructures are built upon pre-established

economic relations between individuals. Moreover as societies (r)evolve, an internal struggle between forces (the state and its institutions) and relations of production (material and social reflections of exchange between individuals and classes) arises, leading to cyclical social revolutions, through which new economic relations manifest (Marx 1859/1970, p.11-12). Marx's understanding of the (re)evolution of history led him to envision an inevitable rise of a classless society at the very last stage of societal conflict. To make such a claim on the future, Marx clearly possessed a conception of a universal human nature that has and will remain at odds with society until this final resolution. We argue that altruism, as the polar opposite of pure self-interest and a dramatic improvement over even enlightened self-interest, is a **necessary condition for the existence of such a society where the means of production are shared and the possibility of exploitation and competition extinguished.**

The Abolition of Morality

For Marx, social superstructure, based in the material conditions of life, dictate ideologies and rationalities down to the individual, and all norms, values and mental activities are conditioned by it. "Men make their own history," writes Marx,

...but they do not make it as they please; they do not make it under self-selected circumstances, but under circumstances existing already, given and transmitted from the past. The tradition of all dead generations weighs like a nightmare on the brains of the living (Marx 1852/2008, p. 15).

For example, both Averroes and Thomas Aquinas were contemporaries and preeminent philosophers of their time. They both shared a vital interest in Aristotelian philosophy, but from each, a different reconciliation between Aristotle and their respective religions, Islam and Christianity, was reached. The social superstructure had conditioned their mode of thought, but did not dictate specifically in which direction their thought should carry them. It is only after too much disharmony within a society that a new form of society is born, and one that is a step closer to a final resolution of conflict, and a final emancipation. In summation, there is a universal human nature, and a relative one shaped by the material conditions of each historical epoch (Karl, Fernbach et al. 1867/1977, p. 759). How and in what form this universal nature expresses itself is discussed below.

Alienation/Dehumanization:

For the first time in history, through capitalism and its fundamental origins in the "enclosure movement" beginning as early as the 12th century, labor power has acquired the essential character of a commodity. It is readily purchased and sold on the market according to the needs of producers and possessed by labor as the only commodity they are capable of selling in the marketplace to subsist (Karl, Fernbach et al. 1867/1977, p. 269). In Marx's earlier writings, this institutionalized and regular transformation of labor (actual exertion) from an individually possessed and expended effort, to a commodity in the form of "labor power" (the potential

exertion bought and sold in the market) has led to a state of alienation and fierce competition. His later writings, while they appear to exclusively and technically address the labor power-commodity and ultimately the actual labor exerted as an exploitable resource used in the generation of profit, should be read with the understanding that there is no "early" vs. "late" Marx. The philosophical underpinnings present in, for example, *The Economic and Philosophical Manuscripts*, ripple prominently through *Kapital* as Marx explores the commodification of labor power in the sphere of circulation and the exploitation of labor within the sphere of production.

The term 'alienation' has the same general meaning, but many different connotations depending upon its use in literature or economics. In the general sense, it describes a state of estrangement from what is humanly natural or in Marxian terms, an alienation from an individual's "species-being". As Erich Fromm comments in "Marx's Conception of Man": "Marx...does not consider capitalism to be the outcome of human nature and the motivation of man in capitalism to be the universal motivation within man". (Fromm 2004, p.11). In Marx's centric view of labor theory, alienation refers to the separation of man from meaningful, subjective personal labor for the sake of being an objective commodity under the mercy of market forces (Fromm 2004, p. 37). Human nature is fundamentally creative but labor within a capitalist framework has creativity stripped from it in the name of efficiency and profit. This is the most immediate form of alienation Marx explores. More generally, as noted in the *Economic and Philosophical*

Manuscripts of 1844, the buying and selling of labor power necessarily creates competition between workers for employment, and competition between producers for profit, giving labor an objective character in the form of a commodity (Marx 1927/1967, p. 19). Aside from being alienated from our "species-being" of humans, which is at least substantially a result of work being fundamentally "uncreative" and routinized, workers are also estranged from others as their social nature is suppressed by a competition systemically necessitated. More fundamentally, workers are estranged from the means of production which Marx viewed as the master form of alienation from which all other forms spring. This latter reality is historically unique to the capitalist era and highlights the degree to which the egoism of capitalism is not an eternal and immutable state of affairs.

The forms of alienation described above begin with alienation from the self which is a fact in both ideological and material spheres. We are not simply denied the creative processes that used to embody work in previous eras of human history. We are also alienated from ourselves in a material sense as the products of our labor that, in a very literal sense, embody "us" are stripped from us by those with ownership of the means of production. This is but one means by which alienation from the means of production becomes the governing form of the capitalist phenomenon of estrangement. It would, to a substantial degree, not be possible were it not for an elitist social structure which placed one individual over another (another form of "person to person" alienation) by virtue of the ownership relation that

regulated their interaction such that the worker is forced to compete with others of the same class in order to gain access to this "dead labor" (Marx's phrase describing Capital) through which subsistence is enabled.

This bilateral material and social relationship between labor and capital is embodied in Marx's words when he noted that in a capitalist society, "[t]he worker produces capital, capital produces him...capitalist and worker taking advantage of each other is shown to be the normal relationship" (Marx 1927/1967, p. 85). The irony of this statement however, is that the Capital made by workers (like history) remanufactures workers in a perverse and destructive manner. We make Capital and are remade by it, but not in a way that we choose. We make capital and, as Marx notes, impoverish ourselves by engaging in the very process we use to seek material enrichment. We seek subsistence and we thereby psychologically and socially alienate ourselves (from ourselves and others) while simultaneously reinforcing the power that Capital has over us by generating surplus labor and profit for the "owning class". We are, in effect, psychologically and socially (and perhaps even materially) suicidal in seeking to live. This is the fundamental internal contradiction that Marx viewed as rippling through capitalism at every turn. By contrast, "[c]ommunism...equates humanism with naturalism, communism as a final resolution of the conflict between man and nature and between man and man" (Marx 1927/1967, p. 99). It was precisely market fundamentalism, or the laissez-faire economics of the time that Marx opposed, and that which is most prevalent today in the new global economy

(Stiglitz, 2004). It is perhaps equally important to recognize that an economic and social environment being governed by "markets" is by no means, if we are to accept human history as evidence, the "natural" state of the world. Perhaps the most vocal proponent of this reality in the 20th century was Karl Polanyi who, in a famous example detailing the Trobriand Islanders, noted that (in spite of their quite detailed exchange mechanisms and long distance trade routes, exhibited an absence of markets and (perhaps even more interesting) an absence of permanent ownership. Further, aside from the aversion to private property, status was afforded in the culture to those who could most efficiently rid themselves of property. In short, altruistic behavior had become the basis of a peculiar sort of social selfishness. (CITE POLANYI). Interestingly, alienation in such a sphere might come about only as a result of becoming a social outcast due to material egoism.

In spite of this and other counterexamples, it has become popular myth that selfishness, and the sorts of alienation that come with it, is a "natural" state of the world. As mentioned above, 'alienation' from society is a condition; the individual is helpless victim of rampant economic forces, where buyer and seller are independent but set against each other (as individuals and classes) in a vicious battle of self-interest. As mentioned above, in Marx's later writings the issue with capitalism was that workers are ultimately reduced to a simple means (living at basic levels of subsistence) to generate surplus (an end over which they have no claim), rather than the

labor product being an end in itself for satisfying a social and personal need (Karl, Fernbach et al. 1867/1977, p.280). This reality relates clearly to the "early" Marxian notions of estrangement and the "primitive accumulation" which was its cause centuries before.

From the analysis presented above, several tentative premises can be drawn from which Marx had deemed self-interest, profit, and capitalism as a whole severely detrimental to individual and collective well-being. This should come as no surprise as human beings are predisposed to exercise their creative will through labor and labor in a capitalist society objectifies personal creative will as a commodity. Therefore, due to the historically determined ownership arrangements over capital, estranged labor necessarily becomes merely a means of sustenance for the worker and a means of profit for the producer.

It therefore stands to reason that, the market based selling and buying of such commodity introduces competition and undermines personal autonomy. Given this, a system of production free of self-interest and monetary gain is key for the return of the individual to his or her natural state.

Though labor theory places creative production in the center, it also concerns human relations primarily. A society based on pure altruism as the antithesis of pure self-interest (and towards the ultimate goal of absolute

equity, fairness and personal autonomy) is a society free of exploitation, profit and personal gain.

Identifying Universal Needs

After considering the diverse viewpoints above, any definition of the nature of human morality is still nowhere within grasp. Sociological studies conducted on the norm of self-interest gives findings that social norms may be more powerful than individual disposition, but it goes no further than that (Miller 1999). Recent psychological studies on subjective well-being (SWB) attempt to shed light on some universal needs across cultures and nations. These studies are typically conducted via extensive ranked surveys on factors presumed to affect well-being with unemployment, income rates, and social capital being typical, but they may be extended to include other subjective factors such as autonomy and respect. The precise methodology will not be discussed here. However, we note that aggregate results of such studies indicate that social capital (meaning networks of trust and reciprocity among individuals) scores higher than income levels on positive well-being and does not exhibit diminishing returns. Moreover, relative income (comparable equity in relation to others in the same social strata and professional capacity) scores higher than absolute income, which tends to suffer diminishing returns around a median level and above, and also (at best) correlates moderately with positive well-being (Biswas-Diener, Diener et al. 2004; Ryan, Deci 2000, Helliwell, Putnam 2004).

These findings indicate the presence of universal needs irrespective of cultures and systems of governance. A governing morality tilted towards equity and social connections, as Marx had stressed, rather than one placing a blind emphasis upon pure material hoarding may be desirable... but such is the limit of SWB studies. They are certainly more exploratory than explanatory. Establishing the independence of effects is also not fully possible. For example, income levels may affect disposition towards social capital. For a more rigorous treatment of universal morality, controlled economic games and neuroscience studies provide alternative frameworks for understanding morality and the roots of altruism.

Economic Games in Neuroeconomics

A number of studies on Economics and Evolutionary Game Theory have used empirical and modeling approaches to assess the possibility of altruism and its dynamics. They are discussed here in light of economic studies, but it should be noted that the similarities with evolutionary studies are often small regarding their principles. Among these, commonly-used experimental designs can be grouped as:

1. Ultimatum Game: Player A (The proposer) is given a certain amount of money. Player A may split the amount with Player B (The respondent) any way he or she pleases. Player B may accept or reject the offer depending on its fairness. If Player B

rejects the offer, Player A surrenders the entire stake back to the researcher. By not accepting the offer made by Player A, this effectively gives Player B the power to punish Player A for unfair offers, at a personal cost, and puts pressure on Player A to make fair offers.

2. Dictator Game: This game is similar in concept to the Ultimatum Game, except Player B has no power to reject an offer. It is purely out of Player A's altruistic motivation to share his or her allotted stake to which ever degree they see fit.

3. Prisoner's Dilemma: Two players are given an option to defect or cooperate simultaneously and anonymously within a limited time frame, upon which a monetary sum is awarded to each player based on mutual or singular acts of cooperation or defection. The game is played a number of pre-specified rounds, known to each player. If both players choose to cooperate, the stake is divided equally. If either player defects and the other cooperates, the defector receives a sum higher than half of the total stake, and the cooperator receives nil. If both players defect, the stake is divided equally, but less than half the total stake for each player. The purpose of PD is to examine direct reciprocity, and retribution. It is strategically advantageous to build trust through successive rounds of mutual cooperation. However, a break in trust by one player for the sake of receiving a higher payoff (one player defects

following mutual cooperation) is likely to invoke a desire for retribution in the cooperating player.

Research from early 2000s onward, spearheaded by Neuroeconomist Ernst Fehr continues to challenge the classical economic concept of *Homo economicus*—a conception of humans as purely self-interested individuals in a hunt for material gain. Fehr proposed a radically different conception that is based on fairness and pure altruism: Strong Reciprocity, or as Fehr defines it “the drive to be kind for those who are kind, and punish those who are unkind”; and also, within that same definition, to make a sacrifice with no immediate or future material gains in sight (Fehr, Fischbacher et al. 2002). In an oft-cited and exemplary study, Fehr et al. (2002) attempted to demonstrate the prevalence of strong reciprocity on a cross-cultural level; Participants from 15 small-scale societies (divided into horticultural, nomadic herding, agricultural, foraging and sedentary societies) were selected to participate in a series of one-shot Ultimatum Games (UG). If the proposers were strictly self-interested they would have to consistently offer the least amount possible of the original stake or none at all, and the same would inversely apply to the respondents (they would consistently accept any offer made no matter how low rather than forego all possible monetary gains. There was a wide variance in mean offers across cultures; on average proposers made offers ranging from 25% to nearly 60% of the original stake. Offers deviating from the calculated mean offers for each cultural group were also frequently rejected (Henrich, Boyd et al. 2001).

A 2006 follow-up study replicated their previous findings with a larger sample base ($n=1762$), and further augmented the study with a Dictator Game, and a three-way interaction UG. The purpose of the study was to probe for a social norm based on punishment to explain cross-cultural variation in mean offers (Henrich et al., 2006). In a three-way interaction UG (TPPG), a third player receives half the stake awarded to Player A. Player A then makes an offer to Player B (which, unlike the two-way UG, he or she has no right to reject that offer). Player C observes the interaction, and may choose to punish Player A by sacrificing part of their own share, to be deducted three-fold from the stake given to Player A. A multivariate regression model was assessed for the standard two-way UG based on the Mean Minimum Acceptable Offer (MAO) as the dependent variable, and age, sex, stake size, income, education, and relative wealth as individual predictors. The entirety of these individual predictors, explained as much as 15.8% of between-population variance, indicating that a significantly large portion of unexplained variance lies outside the applied demographic and economic predictors. Following the same regression method for TPPG, individual predictors were also quite low in terms of their impact on variance (roughly 11%).

Least-squared regression analysis based on MAO in UG as a dependent variable and mean DG offers (after adjustment for shared history via a continental dummy variable, and weighed sample sizes) had a coefficient of 0.31. This provided a better prediction model between degrees of

punishment and altruism compared to aggregate demographic and economic predictors. A possible explanation given by the authors was for the existence of intrinsic altruistic motivation operating outside socioeconomic conditions.

Later research hypothesized that community size relates directly with degrees of third party punishment, given the higher difficulty of maintaining cooperation as a social norm in larger communities where anonymous interactions are more common, and repeated interactions are ephemeral. A strong correlation between the MAOs of TPPG and local community population ($r\text{-squared}=0.54$), and total ethnic population ($r\text{-squared}=0.61$) (Marlowe, Berbesque et al. 2008). In 2010, a third study on the same cultural groups elucidated the roles of Market Integration, Community Size and Religion in altruistic motivation and punishment, through a series of regression models. The same baseline predictors mentioned above (along with community size, world religion and market integration) were modeled against MAOs and mean offers in UG, DG and TPPG. Results indicated that community size correlates positively with MAO in TPPG and UG to a highly significant degree. Market integration and participation in a world religion were able to explain nearly all the variance observed in mean UG, DG and TPPG offers (Henrich, Ensminger et al. 2010). Taken together, these findings indicate that fairness increases in direct proportion with market integration and participation in a world religion, as well as willingness to punish with community size. In effect, larger societies require harsher punishment manifest in mean acceptable offers in UG, DG or TPPG, while fairness in

terms of mean offers across all games covaries to a significant degree with market integration and participation in a world religion. Increase in community size may be directly linked with the enforcement of third party punishment (or rather the manifest social norms) as a necessary condition for maintaining cooperation outside kin based and reciprocal relationships in larger societies, facilitated by large scale market and religious institutions. Intrinsic altruistic motivation may exist, but it is also greatly influenced by the inclusion of social norms associated with larger societies.

Thus, it can be argued that the dogma of economic self-interest in neo-liberal economics is fundamentally flawed and not at all ubiquitous. Altruism for the sake of long-term or short-term reciprocated benefits is excluded in the experimental design of the above mentioned economic games by the incorporation of both anonymity and one-time interactions. However, adopting a dogmatic view of pure, ubiquitous altruism would also be seriously flawed because deviations from the optimal 50/50 fairness thresholds in equity differ on a community-by-community context. Acts of pure selfishness, as low and as ubiquitously punishable as they are, are still consistently revealed in these studies. There is a need for further investigations into the motivation of altruism to explain these deviations from predictions.

The Neural Substrates of Altruism: The fMRI Investigative Method

Functional Magnetic Resonance (fMRI) is a non-invasive tool for measuring local brain activity in response to treatment conditions. The technique exploits the fact that post-synaptic neural depolarization and neurotransmitter action increase regional blood flow. Magnetic distortions in local tissue caused by the slightly paramagnetic deoxyhaemoglobin are captured by an MRI scanner and correlated with increased neuronal activity, after correction for several hemodynamic factors such as cerebral blood flow and volume (Matthews, Jezzard 2004, Logothetis 2003). The technique can, therefore, be used to determine which areas of the brain are stimulated when a subject is faced with a problem, situation, or decision. As a tool, fMRI allows a degree of resolution on intact subjects that is lacking in ablation or neurosurgical approaches. However, it is pertinent to note that by itself data obtained cannot be interpreted unless the area of the brain stimulated has a known (or at least putative) function. To date a number of regions and structures in the brain have been ascribed roles in moral decision-making, reward and pleasure, and problem evaluation and solving. Table 1 summarizes these neurological structure-function relations.

The Neurological Structures of Altruism: Results from fMRI Investigations

Several fMRI imaging studies have focused on economic games and its neural correlates, and have indeed verified several brain regions correlated with intrinsic reward circuitry in response to fairness or inequity. For

example, it was demonstrated that accepting high fairness offers in UGs' (near the 50% mark of the stake) increased activity in several reward-based brain regions in responders, namely, the striatum, VMPC, OFC and midbrain regions. Only the VLPFC and the anterior insula were differentially active when accepting or rejecting unfair offers, respectively. The relationship between VLPFC and the left anterior insula was also inversely proportional in magnitude of activation, suggesting possible negative emotional deregulation through reappraisal (Tabibnia, Satpute et al. 2008).

An earlier study, which focused solely on unfair offers through respondents, found proportionally greater activation in the bilateral insula, ACC and the DLPFC in direct relation to the degree of inequity. Accepting unfair offers required greater activation of the DLPFC and also inversely with greater bilateral insula activation when rejecting unfair offers (Sanfey, Rilling et al. 2003). In both the latter study and Tabibnia et al. (2008) the DLPFC and the right VLPFC were proven to share the same bilateral function in negative emotional processing, with the exception of the DLPFC being more general (cognitive control over neutral and negative emotional processing), and the delineation of function of negative emotional control to the VLPFC according to the valence of the emotional conflict (Golkar, Lonsdorf et al. 2012). The previous findings were confirmed by another study, with the exception that activation of the VLPFC over the DLPFC when accepting unfair offers was also observed (Tabibnia, Satpute et al. 2008).

Given that the ACC/Prefrontal cortex network is consistently activated in negative emotional reappraisal and cognitive control, it can be surmised that accepting unfair offers elicits a negative emotional response that varies from individual to individuals, but one that consistently requires cognitive intervention to override such a negative emotional state.

In an experiment designed to test the effects of cooperation, test subjects participated in a series of simulated maze experiments while under fMRI (Krill, Platek 2012). Test subjects were instructed to complete a simulated maze either in pairs or solo. As pairs, one participant plays the role of an instructor with a map giving simple directions (turn left/right etc.) to the other (blindfolded) participant through an earpiece. Under solo conditions, a single participant was instructed to work their way through the maze without a map, but only through trial and error. Significant activation was found in the OFC and the striatum (specifically the caudate and nucleus accumbens) in cooperative maze tasks versus solo conditions, as well as regions of the parietal lobe, implicating reward-center activation through cooperation. Relative contrast of maze completion in pairs, and under solo conditions revealed greater activation of the OFC and the striatum, thereby implying greater intrinsic reward for effective cooperation (Krill, Platek 2012).

Finally, in the first of its kind, one study depolarized both the right and left DLFPC (right and left-depolarized) using transcranial magnetic stimulation (Hallett 2000) and observed brain functional activity in

respondents participating in an UG (depolarization effectively inhibits activation in targeted brain regions). Results indicated higher differential activation of the right DLPFC in the left-depolarized DLPFC case and the VMPFC in contrast to the right-depolarized case, coupled with higher rejection rates of unfair offers. Neither the right DLPFC nor the VMPFC alone correlated with rejection of unfair offers, thus failure to recruit the right DLPFC resulted in disruption of the VMPFC network where preference-based decision making and evaluation of subjective outcomes take place (Baumgartner, Knoch et al. 2011). The entirety of the ACC/Prefrontal cortex network (including the VLPFC, DLPFC and VMPFC) was consistently activated in all mentioned studies herein and elsewhere for the suppression or reappraisal of negative emotional states. The unique and specific function of each of the VLPFC, DLPFC and the VMPFC remains unclear (Ochsner et al., 2005), but more transcranial magnetic stimulation studies should elucidate the individual network paths in the far or near future. Despite the overlap and ambiguity of the PFC/ACC network, the overall message is clear; inequity or unfairness elicits a negative emotional response that requires cognitive control for suppression or reappraisal.

The Function of the DLPFC

DLPFC activation in the studies above still remains ambiguous, but other studies have tied its function to psychopathological traits. The DLPFC hyperactivation was also specifically implicated in making utilitarian

judgments (Moll, de Oliveira-Souza 2007), violating personal norms (Greene, Nystrom et al. 2004) and executive, rational control over moral judgments as opposed to empathy-based judgments (Brüne, Scheele et al. 2012), as well as voluntary suppression of sadness (Plassmann, O'Doherty et al. 2007). In turn, utilitarian-based moral judgments were correlated with anti-social, Machiavellian and psychopathic traits (Bartels, Pizarro 2011; Koenigs, Kruepke et al. 2012). Other brain regions hypothesized to be involved in dictating sociopathic behavioral traits are the amygdala and the OFC; a dysfunction in either, or the disruption of the OFC-Amygdala limbic network positively correlates with sociopathic behaviour (Craig, Catani et al. 2009).

An early study of the PD game with randomly selected test subjects showed that test subject experienced an activation of the striatum and the OFC when cooperating, and degree of activation of the OFC was positively correlated with frequency of mutual cooperation (Rilling, Gutman et al. 2002). In terms of willingness to cooperate and its associated brain activations, one study attempted to correlate psychopathic scores to differential levels of cooperation. Participants were scored on a psychopathy continuum scale system and played a series of PD games with a computer. To simulate reciprocity, the computer algorithm implemented always reciprocated acts of cooperation, and reciprocated defection in a previous round 67% of the time (if a human player defects, there would be a 67% chance that the computer would also defect the following round). Higher psychopathy score were correlated with higher tendencies for defection

following mutual cooperation. Individuals with higher psychopathy scores exhibited lower OFC activation, inversely correlated higher right ACC/DLPFC activations in making the choice to cooperate. In contrast, less psychopathic individuals exhibited higher ACC/DLPFC when defecting, indicating a normal emotional conflict and cognitive resolution. In terms of outcomes, a player versus computer CD outcome invoked higher amygdala activation in low psychopathy individuals, but no other outcome (DD, CC or DC) showed any differential contrast between low and high psychopathy (Rilling, Glenn et al. 2007).

Thus, perhaps not surprisingly, psychopathy correlates with a repetitive desire to defect, and less regard for positive reward outcome associated with cooperative behavior, due to a dysfunction in reward assessment. It is also interesting to note that no significant activation was found in the VMPFC in association with higher activation of the DLPFC in highly psychopathic individuals, ruling out the possibility of a goal-subjective reward conflict in defecting following mutual cooperation. Results from another study support this hypothesis. When six patients with damage to the VMPFC were engaged in PD, UG, and trust games, all six were willing to part with less as proposers and felt a diminished sense of guilt for being unfair. (Krajbich et al., 2007)

Conclusion and Discussion

The empirical evidence provided by economic games and the neural correlates of altruism suggest that (perhaps thankfully) purely self-interested motivation in humans is rare. Pure altruism or strong reciprocity prevails, even when reciprocity and possible threats of reprisal are absent. This appears to undermine the very basis of market fundamentalism founded on rational self-interest. Simultaneously, these results also give credence to Marx's observations that a society free of greed and repression represents the most legitimate expression of intrinsic and universal morality. Indeed, the results may highlight more fundamental realities that the historical record seems to have willingly cast aside. It is likely that "nature" and the "natural state of man" evolved to encourage cooperation for very practical purposes and that the predatory and pathological (indeed psychopathic) tendencies have come to socially and politically dominate inside an construct artificially created by those with the tendencies and some measure of power. This reality has allowed the "natural" state of cooperation to be replaced by a world in which competition and egoism exist as a temporary survival mechanism patiently waiting for a return to "nature".

Furthermore, the consistent activation of reward mechanism in the brain in cooperating and exhibiting altruism, debunks Kant's rational motivation for altruism. Duty as the rational ideal motivator for altruism cannot be realized, as altruism cannot in turn be realized without a sense of emotional gratification. On the contrary, it would seem that emotional deregulation (through the right DLPFC) and suppression of reward centers

leads to more utilitarian-based decision making. Hyper-activity in the DLPFC, and damage to the VMPFC had already been correlated with sociopathic traits as well. However, activation of the DLPFC-VMPFC limbic system is necessary for rejection of inequity, and thus it would seem that a balance must be struck to preserve the right to equity, and to reject inequitable conditions.

On a cautionary note, it may be that cooperation, punishment and fairness linked herein to intrinsic reward might simply be a function of social norm enforcement or deviation. For instance, a study found that adhering to a norm was directly linked to reinforcement learning (specifically the striatum) (Klucharev, 2009). An earlier study linked social norm adherence to the DLPFC, VLPFC and the OFC (Spitzer et al., 2007). The issue with brain imaging studies is that every study attempts to fit its findings with some pre-existing social cognitive theory. For instance, the matter of emotion and cognition and their relationship is not yet resolved. Do emotions act as the first cause of moral judgment, guide it through preferences and biases, or exist purely as a feedback tool (a conscious afterthought following rational deliberation on anticipated emotion or reward) (Baumeister et al., 2007)? Whatever the case may be, it is certain that emotional processing is intertwined with cognitive reasoning in moral judgment.

There is also the matter of dual-processing. The theory states that two modes of emotional reasoning in moral judgment exist. One mode is

automatic, intuitive and parallel-wise (System-I), and the other slow, deliberate and sequence-wise (System-II) (Evans, 2003). One is impulsive and heuristic, the other reflective and analytic. Through experiential learning, System-II may come to replace System-I as an intuitive, autonomic response to familiar dilemmas (Evans, 2007). The implication of this dual system is the possibility of detecting norm violations and adherence empirically by noting minute reaction times of respondents and proposers in economic games. A possibility not yet explored, but if such possibility is fully realized, the link between reward centers and altruistic behavior may be indefinitely disassociated from or deterministically *linked to* adherence to any specific norm/biases. In the former case, a universal moral compass may be implied, from which a social norm naturally extends and is brought to reality as moral codes. The latter case would prove that only a relative morality exists. One that is continuously molded and shaped in all possible directions by an external moral code, even if such a moral code advocated hatred and mistrust, reward-based learning brain regions would continuously serve to enforce it.

Some limitations regarding the interpretation of neurological data must be acknowledged, nonetheless. Local brain regions are heterogeneous in function, and thus their proposed contextual functioning can only be logically deduced rather than proven indefinitely. Specific brain network and sequence of activations have not yet been fully elucidated nor has context-specific functioning been exhaustively studied. The role of certain proteins on

modulating emotional and rational reasoning such as serotonin (Crockett et al., 2008) dopamine (Bachner-Melman et al., 2005), vasopressin (Ebstein et al., 2011) and oxytocin (De Dreu et al., 2010) and their mutations relation to altruism would be needed to explain variability in behavior. For instance, the catechol-O-methyl transferase Val158Met polymorphism was associated with positive well-being (Wichers et al., 2007) and degree of charity (Reuter et al., 2011). It was known at the time that a single amino acid substitution (Val to Met) as a result of G to A substitution would cause the Methionine variant to catabolize dopamine in the brain four times less the regular rate. A quick scan of the dbSNP page for this polymorphism reveals a minor allele frequency of A=0.389/850 as taken from the 1000 genomes project, with the homozygous A/A forms being much less across all populations. There have been several genome-wide association studies on various personality traits including neuroticism and mood (Luciano et al., 2012) but none have touched upon prosocial behavioral traits. Kin selection theories would also have to be investigated thoroughly to explain altruism outside close-group relations.

Why Does Rational Self-Interest Persist?

Arguments regarding economic self-interest shaping moral decisions are not new and have been applied prominently to the management of natural resources, as an example. Maximization of economic individual benefit is perpetuated because the negative effects of individual actions (e.g., added exploitation of common resources) are shared by all, whereas the positive

effects of such actions financially benefit the individual. This “tragedy of the commons” (Hardin 1968) illustrates a framework incompatible with altruistic behavior if we consider restraint from maximizing economic gains as a manifestation of altruism. Any control over human impulse to maximize gains by overexploitation can only come about from rules put in place by a majority and based on mutual coercion, not self regulation of behavior (Hardin 1968, pp 7-8). In fact, Hardin believed that appeals to conscience in order to regulate overexploitation or damage of natural resources of economic interest were self-defeating in evolutionary, psychological, and philosophical contexts.

Ethical egoism or the so-called “Rational self-interest” has had broad implications on the formulation of capitalism and modern society since the time of Adam Smith onwards (Sen, 1977). In neoclassical economics, individuals in a society are narrowly defined as self-serving “agents” with the sole purpose of maximizing monetary gains (utility or profit), and minimizing losses (disutility). This belief about human nature is a powerful social norm in classically liberal-democratic societies with descriptive and prescriptive effects on the individual. The norm of self-interest is characterized by some social theorists as a “positive feedback loop...between theory and social structure” that ultimately leads the layman to become “a self-interest theorist”; holding the belief that monetary incentives shape behavior of all people despite the general personal divergence of behavior from pure monetary self-interest (Miller 1999). In short, and as referenced above" the

belief in self-interest for its explanatory power of human behavior, and perhaps the existence of an artificial institutional environment founded upon it, has led it to become a self-fulfilling prophecy regardless of actual subjective attitudes and behavior. A plethora of experimental and theoretical studies have been forthcoming since the 1960's discussing the incongruence between the norm of self-interest and actual beliefs and motivation. Prosocial actions are often inhibited in individuals for a number of reasons whether it is fear of exploitation, or appearing as deviants from the norm. For example, there exists a common belief that acting simply out of attitude towards a social cause rather than through having a vested interest would provoke anger or disapproval from those who actually do have a vested interest to act. Prosocial action is also commonly attributed to self-interest in a post-hoc justification by the actual participants.

NEEDS TO BE MORE HERE>....>SEXY CONCLUSION? IS THAT MY JOB?

References

Adam, Smith, Raphael David Daiches, and A. L. Macfie. *The theory of moral sentiments*. Liberty Classics, 1984.

Ashraf, Nava, Colin F. Camerer, and George Loewenstein. "Adam Smith, behavioral economist." *The Journal of Economic Perspectives* 19.3 (2005): 131-145.

Bachner-Melman, Rachel, et al. "Dopaminergic polymorphisms associated with self-report measures of human altruism: a fresh phenotype for the dopamine D4 receptor." *Molecular psychiatry* 10.4 (2005): 333-335.

Baumeister, Roy F., et al. "How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation." *Personality and Social Psychology Review* 11.2 (2007): 167-203. Baumgartner, Thomas, et al. "Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice." *Nature neuroscience* 14.11 (2011): 1468-1474.

Baxter, Mark G., and Elisabeth A. Murray. "The amygdala and reward." *Nature Reviews Neuroscience* 3.7 (2002): 563-573.

Becker, Gary S. "Altruism, egoism, and genetic fitness: Economics and sociobiology." *Journal of economic Literature* 14.3 (1976): 817-826.

Birbaumer, Niels, et al. "Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study." *Archives of general psychiatry* 62.7 (2005): 799

Biswas-Diener, Robert, Ed Diener, and Maya Tamir. "The psychology of subjective well-being." *Daedalus* 133.2 (2004): 18-25.

Blood, Anne J., and Robert J. Zatorre. "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion." *Proceedings of the National Academy of Sciences* 98.20 (2001): 11818-11823.

Brüne, Martin, et al. "Empathy Moderates the Effect of Repetitive Transcranial Magnetic Stimulation of the Right Dorsolateral Prefrontal Cortex on Costly Punishment." *PLOS ONE* 7.9 (2012): e44747.

Cardinal, Rudolf N., et al. "Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex." *Neuroscience and biobehavioral reviews* 26.3 (2002): 321-352.

Craig, Michael C., et al. "Altered connections on the road to psychopathy." *Molecular psychiatry* 14.10 (2009): 946-953.

Crockett, Molly J., et al. "Serotonin modulates behavioral reactions to unfairness." *Science* 320.5884 (2008): 1739-1739.

Damasio, Antonio R., et al. "Subcortical and cortical brain activity during the feeling of self-generated emotions." *Nature neuroscience* 3 (2000): 1049-1056.

Dawes, Christopher T., et al. "Neural basis of egalitarian behavior." *Proceedings of the National Academy of Sciences* 109.17 (2012): 6479-6483.

De Dreu, Carsten KW, et al. "The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans." *Science* 328.5984 (2010): 1408-1411.

Dobb, Maurice, *Studies in the Development of Capitalism*. DATE?

Ebstein, Richard P., et al. "The contributions of oxytocin and vasopressin pathway genes to human behavior." *Hormones and behavior* (2011).

Evans, Jonathan St BT. "In two minds: dual-process accounts of reasoning." *Trends in cognitive sciences* 7.10 (2003): 454-459. Evans, Jonathan St BT. "Dual-processing accounts of reasoning, judgment, and social cognition." *Annu. Rev. Psychol.* 59 (2008): 255-278.

Fehr, Ernst, Urs Fischbacher, and Simon Gächter. "Strong reciprocity, human cooperation, and the enforcement of social norms." *Human nature* 13.1 (2002): 1-25.

Fellows, Lesley K., and Martha J. Farah. "The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se?." *Cerebral Cortex* 17.11 (2007): 2669-2674.

Ferraro, Fabrizio, Jeffrey Pfeffer, and Robert I. Sutton. "Economics Language and Assumptions: How Theories Can Become Self-Fulfilling." *Academy of Management Review* 30.1 (2005): 8-24.

Fromm, Erich. *Marx's concept of man*. Continuum, 2004.

Giddens, Anthony. *Capitalism and modern social theory: An analysis of the writings of Marx, Durkheim and Max Weber*. Cambridge University Press, 1973.

Glenn, Andrea L., et al. "Increased DLPFC activity during moral decision-making in psychopathy." *Molecular Psychiatry* 14.10 (2009): 909-911.

Glimcher, Paul W. *Neuroeconomics: Decision making and the brain*. Academic Press, 2009.

Golkar, Armita, et al. "Distinct Contributions of the Dorsolateral Prefrontal and Orbitofrontal Cortex during Emotion Regulation." *PloS one* 7.11 (2012): e48107.

Gordon, Heather L., Abigail A. Baird, and Alison End. "Functional differences among those high and low on a trait measure of psychopathy." *Biological psychiatry* 56.7 (2004): 516-521.

Gottfried, Jay A., John O'Doherty, and Raymond J. Dolan. "Encoding predictive reward value in human amygdala and orbitofrontal cortex." *Science* 301.5636 (2003): 1104-1107.

Greene, Joshua D., et al. "The neural bases of cognitive conflict and control in moral judgment." *Neuron* 44.2 (2004): 389.

Hallett, Mark. "Transcranial magnetic stimulation and the human brain." *Nature* 406.6792 (2000): 147-150.

Hare, Todd A., Colin F. Camerer, and Antonio Rangel. "Self-control in decision-making involves modulation of the vmPFC valuation system." *Science* 324.5927 (2009): 646-648.

Hare, Todd A., Colin F. Camerer, and Antonio Rangel. "Self-control in decision-making involves modulation of the vmPFC valuation system." *Science* 324.5927 (2009): 646-648.

Hare, Todd A., et al. "Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors." *The Journal of Neuroscience* 28.22 (2008): 5623-5630.

Hare, Todd A., et al. "Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors." *The Journal of Neuroscience* 28.22 (2008): 5623-5630.

Helliwell, John F., and Robert D. Putnam. "The social context of well-being." *PHILOSOPHICAL TRANSACTIONS-ROYAL SOCIETY OF LONDON SERIES B BIOLOGICAL SCIENCES* (2004): 1435-1446.

Henrich, Joseph, et al. "Costly punishment across human societies." *Science* 312.5781 (2006): 1767-1770.

Henrich, Joseph, et al. "In search of homo economicus: behavioral experiments in 15 small-scale societies." *American Economic Review* (2001): 73-78.

Henrich, Joseph, et al. "Markets, religion, community size, and the evolution of fairness and punishment." *science* 327.5972 (2010): 1480-1484.

Hobbes, Thomas. *Leviathan: Reprinted from the Edition of 1651, with an Essay*. Clarendon Press, 1909.

Jost, John T., Mahzarin R. Banaji, and Brian A. Nosek. "A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo." *Political Psychology* 25.6 (2004): 881-919.

Kant, Immanuel. *The moral law: Groundwork of the metaphysics of morals*. Routledge, 2005.

Karl, Marx, David Fernbach, and Ernest Mandel. *Capital*. Penguin, 1977.

Kennerley, Steven W., et al. "Optimal decision making and the anterior cingulate cortex." *Nature neuroscience* 9.7 (2006): 940-947.

Kiehl, Kent A., et al. "Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging." *Biological psychiatry* 50.9 (2001): 677-684.

Klucharev, Vasily, et al. "Reinforcement learning signal predicts social conformity." *Neuron* 61.1 (2009): 140-151.

Knutson, Brian, et al. "Anticipation of increasing monetary reward selectively recruits nucleus accumbens." *J Neurosci* 21.16 (2001): 1-5.

Koenigs, Michael, et al. "Utilitarian moral judgment in psychopathy." *Social Cognitive and Affective Neuroscience* 7.6 (2012): 708-714.

Krajchich, Ian, et al. "Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex." *The Journal of Neuroscience* 29.7 (2009): 2188-2192.

Krill, Austen L., and Steven M. Platek. "Working Together May Be Better: Activation of Reward Centers during a Cooperative Maze Task." *PloS one* 7.2 (2012): e30613.

Kringelbach, Morten L. "The human orbitofrontal cortex: linking reward to hedonic experience." *Nature Reviews Neuroscience* 6.9 (2005): 691-702.

Kringelbach, Morten L., and Edmund T. Rolls. "The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology." *Progress in neurobiology* 72.5 (2004): 341.

Lévesque, Johanne, et al. "Neural circuitry underlying voluntary suppression of sadness." *Biological psychiatry* 53.6 (2003): 502-510.

Lieberman, Matthew D., et al. "Putting feelings into words affect labeling disrupts amygdala activity in response to affective stimuli." *Psychological Science* 18.5 (2007): 421-428.

Logothetis, Nikos K. "The underpinnings of the BOLD functional magnetic resonance imaging signal." *The Journal of Neuroscience* 23.10 (2003): 3963-3971.

Luciano, Michelle, et al. "Genome-wide association uncovers shared genetic effects among personality traits and mood states." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* (2012).

Macpherson, Crawford B. *The political theory of possessive individualism*. Oxford: Clarendon Press, 1962.

Marlowe, Frank W., et al. "More 'altruistic' punishment in larger societies." *Proceedings of the Royal Society B: Biological Sciences* 275.1634 (2008): 587-592.

Marx, Karl, and Friedrich Engels. *Manifesto of the communist party*. CH Kerr & Company, 1906.

Marx, Karl. "A contribution to the critique of political economy." (1970).

Marx, Karl. *Economic and Philosophic Manuscripts of 1844*. Moscow: Progress Publishers, 1967

Marx, Karl. *The 18th Brumaire of Louis Bonaparte*. Wildside Press LLC, 2008.

Matthews, P. M., and P. Jezard. "Functional magnetic resonance imaging." *Journal of Neurology, Neurosurgery & Psychiatry* 75.1 (2004): 6-12.

McClure, Samuel M., Michele K. York, and P. Read Montague. "The neural substrates of reward processing in humans: the modern role of fMRI." *The Neuroscientist* 10.3 (2004): 260-268.

Miller, Dale T. "The norm of self-interest." *American Psychologist; American Psychologist* 54.12 (1999): 1053.

Miller, Dale T., and Rebecca K. Ratner. "The disparity between the actual and assumed power of self-interest." *Journal of personality and social psychology* 74.1 (1998): 53.

Moll, Jorge, and Ricardo de Oliveira-Souza. "Moral judgments, emotions and the utilitarian brain." *Trends in cognitive sciences* 11.8 (2007): 319-321.

Moll, Jorge, et al. "The self as a moral agent: linking the neural bases of social agency and moral sensitivity." *Social Neuroscience* 2.3-4 (2007): 336-352.

Ochsner, Kevin N., and James J. Gross. "The cognitive control of emotion." *Trends in cognitive sciences* 9.5 (2005): 242-249.

Ochsner, Kevin N., et al. "For better or for worse: neural systems supporting the cognitive down-and up-regulation of negative emotion." *Neuroimage* 23.2 (2004): 483-499.

Ochsner, Kevin N., et al. "Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other." *Journal of cognitive neuroscience* 16.10 (2004): 1746-1772.

Ochsner, Kevin N., et al. "Rethinking feelings: An fMRI study of the cognitive regulation of emotion." *Journal of cognitive neuroscience* 14.8 (2002): 1215-1229.

O'Doherty, John, et al. "Abstract reward and punishment representations in the human orbitofrontal cortex." *Nature neuroscience* 4.1 (2001): 95-102.

Paulus, Martin P., and Lawrence R. Frank. "Ventromedial prefrontal cortex activation is critical for preference judgments." *Neuroreport* 14.10 (2003): 1311-1315.

Plassmann, Hilke, John O'Doherty, and Antonio Rangel. "Orbitofrontal cortex encodes willingness to pay in everyday economic transactions." *The Journal of neuroscience* 27.37 (2007): 9984-9988.

Ploghaus, Alexander, et al. "Dissociating pain from its anticipation in the human brain." *Science* 284.5422 (1999): 1979-1981.

Polanyi, Karl. "The Great Transformation" DATE????

Quervain, Dominique J-F. de, et al. "The neural basis of altruistic punishment." *Science* 305.5688 (2004): 1254-1258.

Rachels, James, and Stuart Rachels. *The elements of moral philosophy*. McGraw-Hill, 1986.

Rameson, Lian T., Sylvia A. Morelli, and Matthew D. Lieberman. "The neural correlates of empathy: experience, automaticity, and prosocial behavior." *Journal of Cognitive Neuroscience* 24.1 (2012): 235-245.

Ratner, Rebecca K., and Dale T. Miller. "The norm of self-interest and its effects on social action." *Journal of personality and social psychology* 81.1 (2001): 5.

Reuter, Martin, et al. "Investigating the genetic basis of altruism: the role of the COMT Val158Met polymorphism." *Social cognitive and affective neuroscience* 6.5 (2011): 662-668.

Ridderinkhof, K. Richard, et al. "Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning." *Brain and cognition* 56.2 (2004): 129-140.

Rilling, James K., et al. "A neural basis for social cooperation." *Neuron* 35.2 (2002): 395-405.

Rilling, James K., et al. "Neural correlates of social cooperation and non-cooperation as a function of psychopathy." *Biological psychiatry* 61.11 (2007): 1260-1271.

Rogers, Robert D., et al. "Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex." *The Journal of Neuroscience* 19.20 (1999): 9029-9038.

Ryan, Richard M., and Edward L. Deci. "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American psychologist* 55.1 (2000): 68.

Sanfey, Alan G., et al. "The neural basis of economic decision-making in the ultimatum game." *Science* 300.5626 (2003): 1755-1758.

Sescousse, Guillaume, Jérôme Redouté, and Jean-Claude Dreher. "The architecture of reward value coding in the human orbitofrontal cortex." *The Journal of Neuroscience* 30.39 (2010): 13095-13104.

Singer, Tania, et al. "Empathy for pain involves the affective but not sensory components of pain." *Science* 303.5661 (2004): 1157-1162.

Sen, Amartya K. "Rational fools: A critique of the behavioral foundations of economic theory." *Philosophy & Public Affairs* (1977): 317-344.

Spitzer, Manfred, et al. "The neural signature of social norm compliance." *Neuron* 56.1 (2007): 185-196.

Stiglitz, Joseph E. "Capital-market liberalization, globalization, and the IMF." *Oxford Review of Economic Policy* 20.1 (2004): 57-71.

Tabibnia, Golnaz, Ajay B. Satpute, and Matthew D. Lieberman. "The Sunny Side of Fairness Preference for Fairness Activates Reward Circuitry (and Disregarding Unfairness Activates Self-Control Circuitry)." *Psychological Science* 19.4 (2008): 339-347.

Takahashi, Hidehiko, et al. "Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study." *Neuroimage* 23.3 (2004): 967-974.

Veit, Ralf, et al. "Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans." *Neuroscience letters* 328.3 (2002): 233-236.

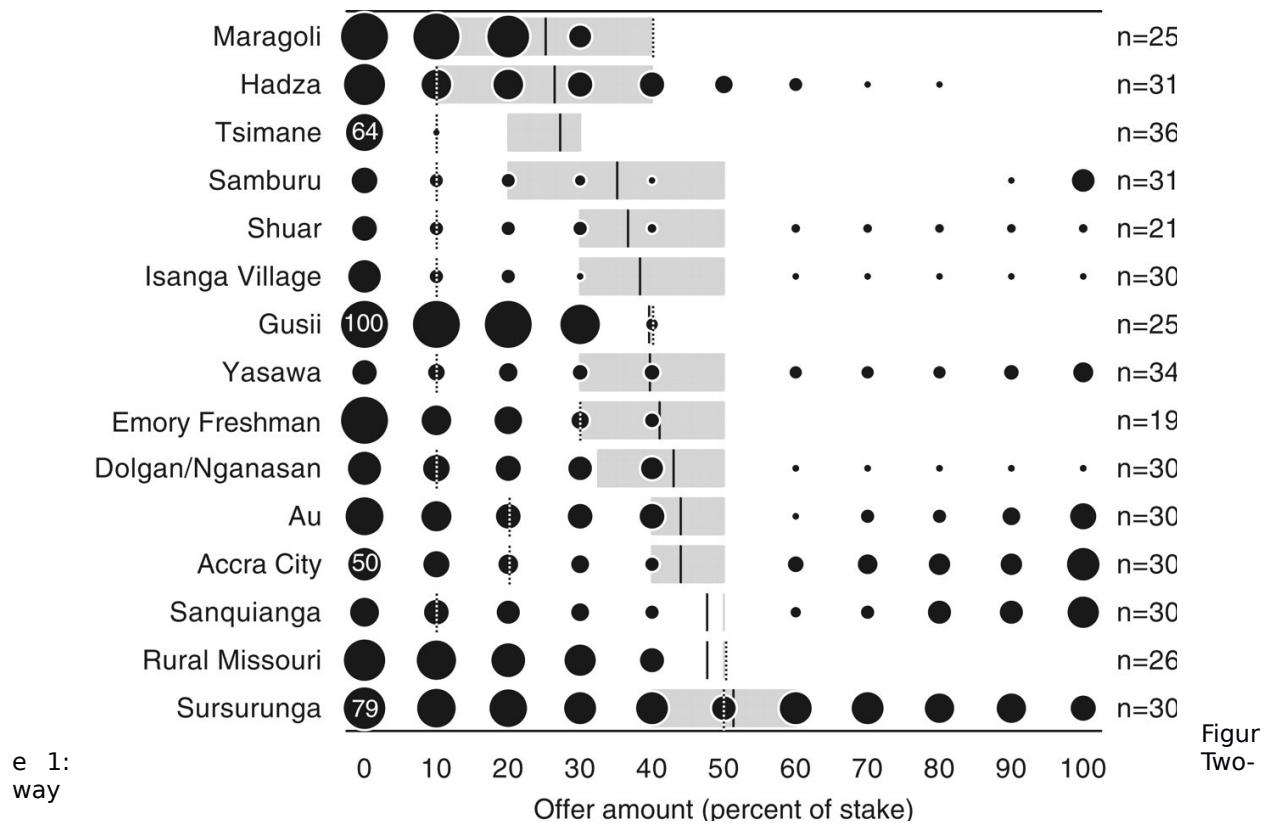
Wallis, Jonathan D., and Earl K. Miller. "Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task." *European Journal of Neuroscience* 18.7 (2003): 2069-2081.

Wichers, Marieke, et al. "The catechol-O-methyl transferase Val158Met polymorphism and experience of reward in the flow of daily life." *Neuropsychopharmacology* 33.13 (2007): 3030-3036.

Wise, Roy A. "Dopamine, learning and motivation." *Nature Reviews Neuroscience* 5.6 (2004): 483-494.

Zalla, Tiziana, et al. "Differential amygdala responses to winning and losing: a functional magnetic resonance imaging study in humans." *European Journal of Neuroscience* 12.5 (2001): 1764-1770.

Tables and figures:



e 1:
way

Figur
Two-

Ultimatum Game results across all study samples. Vertical lines depict the means, while the grey boxes depict interquartile. Circles signify the percentage of offer rejection. The numbers (n) indicate the number of games run. Interestingly enough offers higher than 50% are rejected on an increasing gradient.

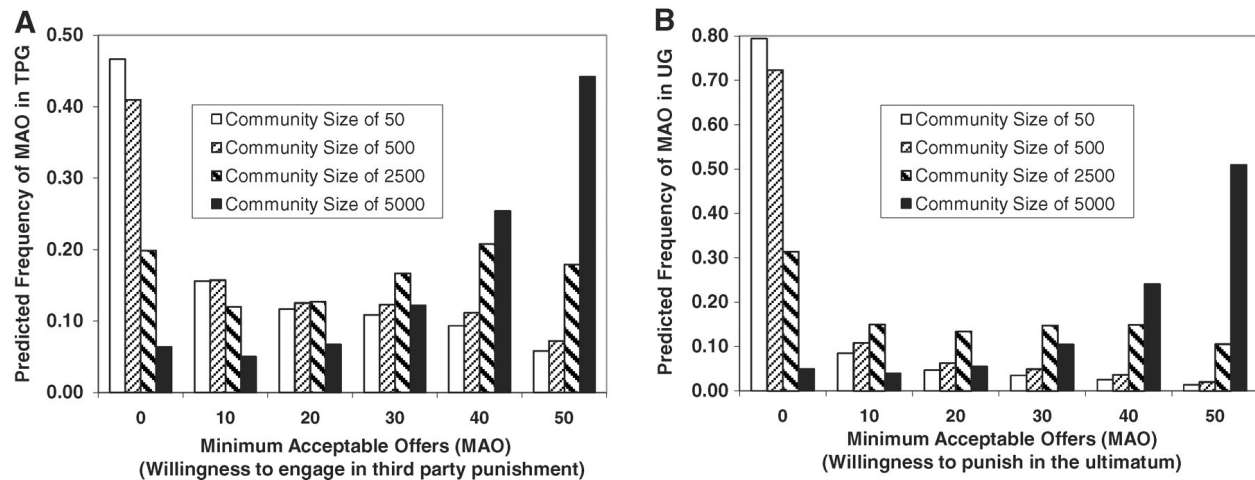


Figure 2: Visual summary of ordered logistic regression model coefficients (based on odds ratio; the increased probability of punishment in the next higher MAO bracket per 100 increase in community size). For example, in UG, an individual in a community size of 5000 has a rough 50% chance to accept an offer of no less than 50% of the stake size, and shows only 5% willingness to accept 5% of the stake size. The same statistic correlated with willingness to punish on behalf of the third party.

Brain Region	Function	References
Orbitofrontal Cortex (OFC)	Focused problem solving	(Kringelbach 2005)
	Learning and memory of positive reward value outcomes	(Kringelbach, Rolls 2004, Hare, O'Doherty et al. 2008, Sescousse, Redouté et al. 2010)
	Punishment evaluation	(O'Doherty, Kringelbach et al. 2001)
	Psychopathy (Hypoactivation)	(Veit, Flor et al. 2002, Birbaumer, Veit et al. 2005)
	Reinforcement/error-adaptive learning	(Kennerley, Walton et al. 2006, Allman, Hakeem et al. 2001)
Anterior Cingulate Cortex (ACC)	Reward-based decision making and reward assessment	(Bush, Vogt et al. 2002)
	Self/other distinction	(Ochsner, Knierim et al. 2004)
Medial Prefrontal Cortex (MPFC)	Pro-social behavior	(Moll, de Oliveira-Souza et al. 2007, Takahashi, Yahata et al. 2004)
	Emotion-based decision making and preference formation	(Paulus, Frank 2003, Fellows, Farah 2007)
Ventromedial Prefrontal Cortex (VMPC)	Reward (pleasure) and Motivation	(Blood, Zatorre 2001)
	Evaluation of (normative) goods	(Moll, Krueger et al. 2006, Hare, Camerer et al. 2010)
	Evaluation of goal-directed decision	(Hare, Camerer et al. 2009)
	Evaluation of predicted reward	(Hare, O'Doherty et al. 2008, Plassmann, O'Doherty et al. 2007)
Medio-orbital Prefrontal Cortex (MOPC)	Reward-based associative learning	(Ridderinkhof, van den Wildenberg et al. 2004)
Ventrolateral Prefrontal Cortex	Negative emotional reappraisal for the purpose of deregulation	(Ochsner, Bunge et al. 2002, Ochsner, Ray

(VLPC)			et al. 2004, Ochsner, Gross 2005)
		Reward and Motivation	(Gottfried, O'Doherty et al. 2003, Knutson, Adams et al. 2001)
Striatum (Caudate and Nucleus Accumbens)		Dopaminogenic injection target	(Glimcher 2009)
		Error evaluation in outcomes from predicted reward	(Hare, O'Doherty et al. 2008)
		Reward-based learning	(Baxter, Murray 2002, Gottfried, O'Doherty et al. 2003)
Amygdala		Reward valence assessment	(Arana, Parkinson et al. 2003)
		Psychopathy (Hypoactivation)	(Kiehl, Smith et al. 2001, Gordon, Baird et al. 2004)
		Neutral emotional reappraisal	(Lévesque, Eugène et al. 2003, Golkar, Lonsdorf et al. 2012)
Dorsal Prefrontal (DLPFC)	Lateral Cortex	Reward value evaluation	(Wallis, Miller 2003)
		Psychopathy (Hyperactivation)	(Müller, Sommer et al. 2003, Kiehl, Smith et al. 2001, Gordon, Baird et al. 2004)
		Negative emotional processing	(Damasio, Grabowski et al. 2000)
Bilateral Insula	anterior	Emphatic pain and distress	(Singer, Seymour et al. 2004, Ploghaus, Tracey et al. 1999)

