Journal of Economic Behavior & Organization 77 (2011) 4–13

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Sympathy with Adam Smith and reflexions on self

David Haig*

Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 27 May 2009
Received in revised form
24 December 2009
Accepted 10 January 2010
Available online 21 September 2010

Keywords:
Reciprocity
Sympathy
Moral sentiments
Evolution
Adam Smith
Self

ABSTRACT

Adam Smith's account of the moral sentiments resonates with modern themes in evolutionary biology. His distinction between our reasons and the reasons for these reasons recalls the evolutionary biologist's emphasis on different levels of causal explanation. In this view, the proximate goals of our psychological motivations are different in kind from the ultimate reasons why we have evolved these motivations. Sympathy was central to Smith's account of the moral sentiments and he discussed two principal forms of sympathy. Second-person sympathy is putting ourselves in another person's situation to see the world from their perspective. Third-person sympathy is viewing ourselves from the perspective of an impartial observer. In recent discussions of the evolution of cooperation, second-person sympathy facilitates cooperation via direct reciprocity, *I behave well by you so that you will behave well by me*, whereas third-person sympathy facilitates cooperation via indirect reciprocity, *I behave well by you so that others will behave well by me*.

1. Introduction

This year marks the 250th anniversary of *The Theory of Moral Sentiments* and the 150th anniversary of *On the Origin of Species by Means of Natural Selection*. The essay you are reading contains the reflections of one 21st-century Darwinist on the passions, reason, and morality, inspired by his re-reading Adam Smith's 18th-century masterpiece. Smith centered his account of the moral sentiments on the concept of sympathy: we understand others by putting ourselves in their situation; and we judge our own conduct by viewing ourselves from the perspective of an impartial spectator. Our faculties of sympathy are both reflexive, automatic responses beyond the control of our wills, and reflective, reasoned contemplation of others and our relations to them. Smith's prose matches his subject, adopting multiple perspectives and switching voices, at times intimate and passionate, and at others distant and reserved. There is a rhythm to his prose that resonates in the mind of the reader and there is, to my mind, a playful seriousness that invites a serious playfulness in reply. *The Theory of Moral Sentiments* is not only a work on sympathy but also a work that evokes a sympathetic response in the reader.

My essay attempts an explication of the moral sentiments that melds the insights of Darwin and Smith. It has three mains sections. Section 2 discusses the multiplicity of guides to individual action under the broad categories of instinct, reason, and culture, and discusses different kinds of answers to the question why we act the way we do. Section 3 explores different kinds of reflections (and reflexions) back upon our internal self image. Section 4 uses the arguments of the preceding sections to discuss the nature of our moral faculties. Our moral choices are viewed as emerging from a nexus of conflicting agendas of different entities with different ends. I will suggest that a locus of moral responsibility, and a sense of self itself, emerges as we flip back and forth between our own perspective and the perspectives of others, and as we attempt to reconcile and adjudicate among the different springs of internal action.

* Tel.: +1 617 496 5152; fax: +1 617 495 5667.
E-mail address: dhaig@oeb.harvard.edu.

0167-2681/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2010.09.002

In thinking about sympathy, my mind kept returning to the metaphor of mirrors reflecting mirrors: we see ourselves through others' eyes who see themselves through our eyes. In keeping with this theme of reflection, and reflections upon reflections, I have attempted to give my essay a recursive structure in which the text constantly reflects back upon itself. As a work of sympathy with Adam Smith, I have not attempted to achieve complete clarity in the text, nor within myself, about when I speak in Smith's voice and when I speak in my own. Such ambiguity seems fitting when discussing a topic that blurs the boundaries between individuals and their not-so-distinct points of view. My views depart most from Smith's treatment of sympathy when I allude to the usefulness of sympathy in manipulating and exploiting others for selfish ends. Perhaps Smith did not consider that instrumental uses of sympathy came under the purview of the *moral* sentiments or perhaps he had more faith than I in the beneficence of creation.

2. Teleology

> "In every part of the universe we observe means adjusted with the nicest artifice to the ends which they are intended to produce; and in the mechanism of a plant, or animal body, admire how every thing is contrived for advancing the two great purposes of nature, the support of the individual, and the propagation of the species. . . . But though in accounting for the operations of bodies, we never fail to distinguish in this manner the efficient from the final cause, in accounting for those of the mind we are very apt to confound these two different things with one another. When by natural principles we are led to advance those ends, which a refined and enlightened reason would recommend to us, we are very apt to impute to that reason, as to their efficient cause, the sentiments and actions by which we advance those ends, and to imagine that to be the wisdom of man, which is in reality the wisdom of God." (Smith, 1976, p. 87)

Smith recognizes two levels of causal explanation in this passage: our reasons and the reasons for these reasons. The specific question to which he alludes is the source of our approbation of the punishment of individuals who violate moral laws. In Smith's opinion, we punish, and approve of punishment, out of our indignation against the offender, not because of a reasoned consideration of the value of punishment for the proper ordering of society. Our indignation, however, has been contrived as an efficient means of advancing the latter end. We act out of passion, but the preservation of society is the reason why we have been endowed with this passion.

Smith's grounding of final causes in God's wisdom could be considered an orthodox appeal to natural theology, but his ontological stance on the nature of final causes is not altogether clear. When he wants to, Smith writes clearly, and his lack of clarity on teleological questions is, I suspect, deliberate. A hundred years later, Darwin provided a naturalistic account of the appearance of purpose in nature: spontaneously-arising variation modifies the properties of organisms; some of these modifications benefit the organism in its struggle for existence and these modifications are thereby perpetuated in the organism's offspring; thus, an *effect* of a modification of the organism is a *cause* of that modification appearing in subsequent generations.

Darwin's understanding of the hereditary material was inchoate. He would have accepted support of the individual, and even the tribe, as the 'goal' of the adaptive process, but many now prefer to view the genetic material itself, rather than the individual or group, as the beneficiary of the fruits of natural selection (Dawkins, 1982). This, however, is a subject of ongoing debate in the philosophy of biology, with most of the polemic heat concerning semantic rather than substantive issues. From a gene-centric perspective, a gene's functions (or purpose) are those of its *phenotypic effects* that have a *causal role* in the gene being preserved and proliferated.

The recognition of different levels of explanation is familiar to evolutionary biologists (Mayr, 1961; Tinbergen, 1963). Consider a serial philanderer who copulates with multiple women by making false promises of commitment. He does not copulate to pass on his genes, but to experience sexual pleasure at little personal cost. The sexual gratification he receives after each successful seduction serves to reinforce the seductive behaviors (intrapersonal recursion). But, the system of sexual desire, seduction, gratification, and reinforcement exists, in part, because he had ancestors who passed on their genes because they consummated their desires by seduction (evolutionary recursion). The philanderer does not copulate to transmit his genes, rather to experience sexual pleasure, but copulation is pleasurable because the promise of pleasure has been the means whereby our ancestors were induced to copulate.[1]

Two levels of *teleology* can be recognized in this example. Sexual desire, and the behaviors it motivates, is the means that achieves the end of sexual pleasure for the philanderer, and the means that achieves the end of open-ended replication for his genes. That psychological motivation and evolutionary function are not the same thing is clearly illustrated by the observation that the philanderer prefers that copulation does not result in conception whereas his genes 'prefer' that it does.

[1] Mayr (1961) distinguished proximate explanations (mechanisms; How?) from ultimate explanations (adaptive function; Why?). Tinbergen (1963) recognized four kinds of explanation: physical causation, survival value, evolutionary history, and ontogeny. I prefer to treat psychological motivation as a fifth kind of explanation, complementary to the others, rather than as a special kind of proximate mechanism or physical cause. When I wish to understand *why* you behaved as you did, I am usually asking a question about the *telos* of your psychological motivations. When I wish to persuade you to do something you might not otherwise do, I am interested in *how* your psychological motivations can be used as a means to my ends.

The philanderer may use a condom and thwart his genes' ends. Or, his genes may promote 'irresponsible' behavior that conceives an *unwanted* pregnancy.[2]

A failure to distinguish evolutionary function from psychological motivation bedevils many debates about our moral nature. Are we ultimately selfish or do we genuinely care for others? Both could be right. Benevolent motivations would not cease to be benevolent if they were shown to be adaptations of selfish genes.

2.1. Instinct

Fitness is the *telos* of our genetic adaptations, but each passion also has a proximate *telos* toward which it cajoles us to action. Hunger has the goal of food; thirst, the goal of water; lust, the goal of sexual gratification; and the desire for breath, the goal of oxygen in the life-saving gasp of a drowning man. Our social desires have social goals; to be accepted, to be admired, to be loved, to be feared, and to be revenged. For want of a better term, I will refer to the proximate *telos* of each passion as its utility.[3]

Utilities are correlated with fitness, but fitness and utility are not the same thing. Children, not genes, are the proximate goal of the desire to have children, a *telos* that is sometimes attained by adoption or ovum donation. Passions may misfire and fail to enhance fitness, but, from our personal perspective, happiness is happiness whether or not our genes benefit. Genes have no 'preference' whether fitness is achieved through our misery or our happiness, through our hatred or our love. But we may have distinct preferences among the alternative paths urged by our passions.

Our instincts can be used as means to others' ends. This is particularly clear when an advertiser uses our hopes and our fears to enrich himself and anonymous shareholders, but the same is true of all forms of persuasion. When the Catholic church persuades a young man to enter the priesthood and forsake genetic reproduction for a religious calling, his motivations are used to serve the ends of a cultural tradition. Nevertheless, the priest may attain the proximate ends of at least some of his passions if he leads a happy and contented life; and he may serve the ends of his moral code as well as the ends of believers' genes if he persuades his parishioners to have larger families.

Genes 'learn' the hard way, by the selective elimination of the less fit. But one of the most useful lessons 'learned' by genes is how to construct organisms that learn from experience and thus compensate for the limitations of innate mechanisms. We learn from the mismatch between our projected and actual performance; from the repetitive performance we call practice; from the mock performance we call play; and from the virtual performance we call reasoning to consequences. And we learn from the performance of others. Our faculty of imitation allows the costs of learning to be shared among individuals and expertise to accumulate over generations. By these means, we come to embody more adaptive information than is present in our genome. We are rational and cultural beings, instinctively. But reason and culture can employ our passions in pursuit of ends other than fitness.

2.2. Reason

Reason encompasses innate problem-solving mechanisms that may be conscious or unconscious and that can be employed in the pursuit of multiple ends. Reason can respond to novel features of our current environment, for which there has been no past natural selection, and can take advantage of small differences in expected return that constitute too weak a selective force for the evolution of hard-wired adaptations.

Reason is both the slave and the governor of the passions. As a slave, reason is employed to find effective ways to further the objectives of the passions. In the process, reason may identify subsidiary goals on the path to satisfying a desire. If I am thirsty, I need to find a safe route into a ravine or a quick way to improvise a bucket. In this manner, reason modifies and gives specificity to the goals of our passions. We have multiple passions because fitness is achieved in different ways in different contexts. Actions that enhanced fitness in one context sometimes will have been disastrous in another context. Therefore, reason, as a governor, has evolved to reconcile and adjudicate among the passions in the context of a specific life, in a specific environment, and a specific culture.

The *telos* of reason – the 'utility function' that compares and aggregates the preferences of different passions – remains for me something nebulous and ill-defined. The theory of social choice (Arrow, 1963), rather than of rational choice, may provide a better model of what goes on within the mind if reason cannot make intrapersonal comparisons of utility when aggregating preferences of different agents within the self. And this theory suggests that basic precepts of normative rationality may be violated if cardinal utilities cannot be compared (Haig, 2006b). Introspection fails to reveal clear criteria by which I resolve internal dilemmas. My decisions are unstable and seem to aim at a *reasonably* happy life while juggling duties and obligations to other individuals, all in the context of conflicting cultural suggestions on how best to balance the passions to achieve a 'good life' (whatever that may be).

[2] Moral codes differ on whether contraception compounds or mitigates the sin of fornication, and on whether the pursuit of pleasure is permissible without the possibility of procreation.

[3] More utility is preferable to less utility for each passion, but I do not wish to presuppose that the utilities of different passions are commensurable or representable on a single scale.

I will leave such foundational questions to one side, assume that reason recommends an action opposed to the blandishments of some passion, and ask by what principles such an internal conflict could be resolved. The conflict between passion and reason is often posed as a problem of self-command or strength of will. From a genetic perspective, this is a conflict between two adaptations. Passion summarizes the wisdom of past natural selection and recommends actions that have promoted fitness in similar situations in the past, whereas reason responds to specific features of the current context and can potentially recognize whether emotion recommends actions that are inappropriate and self-defeating, in this particular instance. But reason too is fallible and its choices can be influenced by the persuasion of other actors who may not have our genetic interests at heart. Neither reason nor passion can be guaranteed to give better advice in any particular case. Strength of will is likely subject to stabilizing selection: too weak a will, and passion will have its way on many occasions when reason is wiser; too strong a will, and reason will ignore too often the conservative wisdom of passion.

Reason can be a passion of its own. Solving puzzles is rewarding in itself. The desire to make sense of the world is a psychological motivation that is consummated in the pleasure of understanding and the thrill of discovery. We never know what knowledge may be useful to achieve future goals, and the solutions we discover to life's conundrums are 'intellectual property' that we can share or trade with others for our advantage.

2.3. Culture

We learn from the experience and reasoning of others and thereby reduce the costs of individual trial-and-error learning. We exercise discrimination in what we learn and tend to emulate the behavior and opinions of those who have achieved, or seem to have achieved, the things we desire in life. Our teachers themselves had grandteachers who themselves had great-grandteachers, with discrimination and recombination exercised at each link in this chain of transmission. By this cumulative process, the ideas which are culturally transmitted evolve to become ever more attractive to learners. Cultural traditions acquire adaptations that promote their own transmission, and thereby possess their own teleology. Culture has a tendency to enhance the utility and fitness of its carriers because we adopt cultural items that appeal to psychological motivations that have been shaped by natural selection. But, culture evolves much more rapidly than the passions and by processes that need not promote either fitness or utility (Haig, 2006a; Richerson and Boyd, 2005). And thus, our psychological motivations become partially untethered from the moorings of their genetic function.

3. Sympathy and the proliferation of selves

> "As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation." (Smith, 1976, p. 9)

Sympathy – the vicarious experience of others' actions, emotions, tastes, and reasons – is central to Adam Smith's account of the moral sentiments. In preparing this essay, I have entered into sympathy with his thought and his rhetoric, and can see traces of his cadence and his style creeping into my thought and my prose. By identifying with him, I have developed affection for the long-dead philosopher, and wish him well, but benevolence is not a necessary part of my definition of sympathy. We also sympathize with others the better to use them as means for our ends, and to avoid being exploited as means for their ends.[4]

Mirror neurons are commonly interpreted as the neurological basis for our sense of sympathy (Gallese, 2007; Molnar-Szakacs, 2011; Fogassi, 2011), but I am no neurologist, and, in emulation of Smith, I will write in generalities and abstractions. Below, I recognize three levels of sympathy. First-person sympathy involves identification with our own self-image and is the scaffold upon which we construct images of other selves. Second-person sympathy involves experiencing the viewpoint of a person with whom we are directly interacting. Third-person sympathy involves viewing our own conduct from the perspective of an impartial observer.[5]

3.1. First-person sympathy

We possess a first-person mental image of the current position of our limbs and other body-parts that is crucial in planning motor actions. Integral to this self-image is perceptual feedback that allows future plans to be modified by past performance. When we miss a target with a thrown projectile, we adjust our posture and point of release at our next attempt. The overlapping representation of action and perception of action allows us to learn from experience and to perfect our aim, by bringing the predicted perception of action into register with the actual perception of action.

Our mental self-image keeps track of many aspects of our physical and emotional state besides the position of our limbs, helping us to adjust, understand, and predict our own behavior. We enter into sympathy with ourselves (first-level recursion) and 'share' in our own hopes and fears. Just as each of us can improve, if not perfect, our jump-shot through self-

[4] Notice how my prose slipped from the first-person singular, in talking of *my* benevolence, to the first-person plural, in discussing *our* less attractive motives. By so doing, I attempted to engage your sympathy to temper your judgment of my sins.

[5] Smith discusses second-person sympathy in parts I and II of *TMS*. His discussion of third-person sympathy begins in part III.

directed effort, we each have a limited ability to set ourselves moral goals and become more like the person we would wish to be.

3.2. Second-person sympathy and direct reciprocity

> "When we see a stroke aimed and just ready to fall upon the leg or arm of another person, we naturally shrink and draw back our own leg or our own arm; . . . The mob, when they are gazing at a dancer on the slack rope, naturally writhe and twist and balance their own bodies, as they see him do, and as they feel that they themselves must do if in his situation." (Smith, 1976, p. 10)

We use our image of our self as the template on which we construct images of other selves, and are thereby better able to understand and predict their actions and to learn from their experience. When we watch a television chef whisk an egg, we utilize her experience, and the experience of her teachers, to improve our souffle, in part by direct bodily simulation of her movements. Bodily sympathy is thus one of the essential components of our faculty of culture. Not only does our second-person image of another allow us to learn from her experience but it also allows us to interpret her feelings, and predict her actions, for our benefit, whether we use this knowledge for her good or her ill.

My second-person image of you incorporates my perception of your current state shaped by my memories of your past behavior. This representation helps me to predict your future behavior in our current relations. Do I *feel* you are somebody who can be trusted, or somebody it would be foolish to trust? Do I *feel* you are somebody who can be exploited without fear of reprisal, or somebody it would be dangerous to cross?

Second-person sympathy is simulated experience not actual experience, and images constructed by the eye of sympathy may be distorted, or out of focus, in a manner analogous to imperfections of the images constructed by our outer eyes. Natural selection favors increased acuity of both inner and outer vision, but this does not preclude occasional sympathetic myopia or cataracts of compassion. My first-person model of myself may be a poor prototype for constructing my second-person model of you because our characters differ, because of failures of my imagination, or because my self-image is itself distorted.

Although our first-person image of ourself is the template for our second-person images of other selves, we are not restricted to viewing others as identical to ourself. We can learn their idiosyncrasies and model their behavior as somewhat different from our own: a skilled boxer anticipates and deflects the punches of both left-handed and right-handed opponents; a skilled philanderer embodies knowledge of female nature, and of individual women, the better to compromise their virtue.

It is relatively simple to adjust for the lesser abilities of others in our sympathetic appreciation of their situation, but difficult, if not impossible, to simulate abilities we do not possess ourselves. A chess master can lure an average player into a trap by viewing the board through his opponent's eyes, but the average player flounders in his attempt to anticipate and block the moves of the master. Natural selection thus favors an escalation in our abilities, including our sympathetic abilities, the better to manipulate others and to avoid being manipulated by them. But, we are less likely to trust those with whom we cannot sympathize and thereby find less predictable. Natural selection may also dampen the development of unusual abilities because of our need to live and be accepted in society.[6]

3.3. Reflected first-person sympathy

Second-person sympathy creates recursive complexity. My second-person image of you includes my simulation of your second-person image of me, and I project this *reflected image* onto my first-person image of myself (second-level recursion). If I feel you feel warmly to me, I feel more warmly to myself. If I feel you find me unattractive, I feel less attractive. I become more aware of myself by becoming more aware of you, and my first-person sympathy is thereby enriched. However, if I have a distorted image of myself, my image of your image of me reflects the distorted image back upon itself, further distorting the original image. If I project a deceptive image of myself to others, I risk becoming a victim of the same deception via the reflected image of myself. Feed-back from reflected self-images must be appropriately damped if our self-image is not to develop extreme distortions or undergo wild fluctuations.[7]

The reflected image of myself provides an answer to the questions: Do I *feel* you feel I am somebody who can be trusted? Do I *feel* you feel I am somebody who can be exploited? In principle, my second-person image of you also includes my simulation of your simulation of my second-person image of you (third-level recursion). This image of you, in the reflected image of me, allows me to answer the question: Do I *feel* you feel I feel you are somebody who can be trusted? My difficulty in mentally parsing higher terms of the series indicates cognitive limits on my depth of recursion that may, in turn, reflect the limited adaptive value of deep introspective self-awareness. Perhaps you find it easy. But then, this is something about you I cannot understand, beyond that you tell me it is so.

The fact that you form a second-person image of me that influences how we interact provides me with an incentive to manage that image. As a result, my first-person representation of myself needs to keep track of the things that I present to

[6] Second-person sympathy can be either reflexive or reflective. When I watch a dancer on a slack rope my body *naturally* writhes and twists as I feel I *must* do in his situation, but when I enter into the mind of an opponent at chess, I attempt to match my reason and self-command to his.

[7] One might refer to such aberrations as sym*pathology*.

your perception — a *projected self* that I intend should mold your second-person representation of me — and the things I hide from your perception — a *private self* that I do not wish to influence that image. As a further consequence, my second-person representation of you will include an estimate of your private self, constructed from things you have unintentionally given away and from guesses about what you hide from me, based on inferences from what I hide from you. Perhaps there are feelings or thoughts I hide from myself, that are excluded even from the image of my private self, the better to hide them from you (Trivers, 2000). And, perhaps, in some cases, we both benefit from some motivations not forming a part of our reflected self-images.

We construct different projected and private selves for different others, one for our mother, one for our spouse, and potentially one for each individual with whom we interact. But in practice, we develop a catch-all projected self, a *public self*, that we present to most others. This consistent self minimizes both the cognitive load of keeping track of a proliferation of self-presentations and the risk of being detected in duplicity by third-party observers. The larger the mismatch between our private and public self, the greater the risk of detection, and the less useful the feedback we receive from our reflected self-images. Integrity is born out of prudence.

3.4. Third-person sympathy and indirect reciprocity

> "Though it may be true, therefore, that every individual, in his own breast, naturally prefers himself to all mankind, yet he dares not look mankind in the face, and avow that he acts according to this principle." (Smith, 1976, p. 83)

A particularly important contribution to our second-person images of others was alluded to at the end of the previous section. You and I may bring to our current interaction second-person representations of each other that have been shaped by observation or report of each other's past interactions with third parties. Moreover, a third-party observer of our current interaction may form a second-person representation of me that will shape his behavior in future interactions with me, for my good or my ill, and that he may share with parties unknown. Therefore, in my interaction with you, I need to maintain and monitor a second-person representation of an observer that includes his second-person representation of me. I must see myself through his eyes and sympathize with his approbation or disapprobation of my character. I must mold my public self to his judgment. Unlike previous second-person images, this representation of someone who might possibly be watching is not a model of a specific individual, but of a generic observer. Its most salient feature, prominent in an otherwise nondescript image, is the reflected image of ourselves as we might appear to others.

We construct different third-party images to represent different kinds of observers to be used in different social contexts. The old woman crying at the Obama inaugural felt a hint of her own acceptance in the President's acceptance by the nation, she re-experienced her own pain of rejection, and perhaps adjusted, a little, the reflected image of herself in her representation of a white observer.

Man was made to live in society. He procures great benefit from interaction with his fellows and would sustain great cost if he were to lose their esteem and be excluded from their company. Although he may gain short-term benefits from exploiting others, he can rarely be sure that no one is watching. The potential future costs from damage to his reputation, and his reduced ability to enjoy the benefits of society, will often dwarf any small advantage he might gain from misbehavior in the present.[8] Thus, in man's mental model of the judgment of a third-party observer, we have arrived at a close relative of Adam Smith's 'man within the breast', the impartial spectator who judges our conduct. Direct reciprocity — I behave fairly by you, so that you behave fairly by me — now shares the stage with indirect reciprocity — I behave fairly by you, so that others will behave fairly by me (Alexander, 1987; Nowak and Sigmund, 1998).

3.5. On tensions between second- and third-person sympathy

Second-person sympathy involves engagement with the perspective of a particular other whereas third-person sympathy involves engagement with the perspective of a generic other. And these two others may view differently how I should act in a particular situation and present me with different reflected images of those actions.[9]

A friend is someone with whom we develop intense second-person sympathy and with whom we partially disengage the third-person perspective. When a friend involves me, without asking, in a conspiracy against the common good, how should I reconcile my moral obligations to friendship with my moral obligations to society? At what point does my desire to remain in good standing with my friend give way to my fear of losing standing in society? And what does my dilemma tell me about my friend? Has he implicated me because he has a faulty second-person image of me (a lack of sympathy for the man within my breast), or because he has an accurate image and knows I will bend? My friend possibly views the

[8] "Nature, when she formed man for society, endowed him with an original desire to please, and an original aversion to offend his brethren" (Smith, 1976, p. 116). In this essay, I use 'man' and masculine pronouns to refer to both sexes. This was the convention in Smith's day but shows a certain lack of sympathy with female readers. I made this choice knowingly, to accentuate my sympathy with Smith, but with anxiety about how it would affect my second-person representation in your minds.

[9] Held (2006) argues that moral philosophy has over-emphasized the claims of third-person detachment and defends an 'ethics of care' based on second-person engagement with particular others. Reddy (2008) seeks to redress a similar imbalance between third-person inference and second-person engagement in theories of how infants come to understand other minds.

opportunity for mutual gain that he offers me to be an expression of the value he places on our friendship and a means of strengthening and deepening that relationship. From his perspective, my moral qualms could be a sign that I care more about abstract principles, and the judgment of others, than I care about him.

A person who is always looking over his shoulder, and considering how he appears to the world, can seem cold and detached. In our most intimate relations, we want a partner who is attentive to us, not the rest of humanity. In a pinch, I want someone who loves *me* and values my interests above others'.

3.6. Reflections on rhetoric and belles lettres

The sole reader of the early drafts of this essay was the writer. His first attempts did not convince and were sent back for revision. In the process of reading his words, and observing his struggles to persuade, my ideas took form and, in a resonance between reader and writer, I came to know my own mind. I, the author, also had a second person in mind. It was you, dear reader. If I have succeeded in reading through your eyes, then perhaps you will see through mine. You will have entered into sympathy with my thought, just as I have entered into sympathy with Smith's. Modern academic writers must also consider a third person; a reader who observes the second-person transaction between author and reader from one side, who does not directly engage with the author's message but with the tools and tricks that the text uses to achieve its effects. What creates passionate engagement with the second-person reader of the author may not always satisfy the reasoned judgment of the third-person reader of the text (Brown, 1994).

4. Morality

Morality is not one thing but rather an imperfect amalgam of instinctive, rational, and cultural elements. As with all things evolutionary, the relations among these elements are recursive. Culture is shaped by instinct, and instinct by culture. Reasoned choices facilitate the evolution of innate behaviors that achieve the same end without reason (Baldwin, 1896). Reason selects cultural items that best serve our ends, and culture is thereby changed, but reason's choices are constrained by cultural norms. Morality is both deeply personal and highly social, something that emerges from within and is imposed from without.

4.1. Instinctive elements

Many aspects of morality are instinctive, modified by personal experience, reason, and culture. Our basic repertoire of emotional responses is innate: including the resentment and indignation we feel toward the selfishness of others, the gratitude we feel toward their generosity, and the guilt or shame we feel for our own trespasses. But our emotional repertoire also includes (among other unamiable passions): the envy, jealousy, and even hatred, we feel toward those who possess something we want; and the desire for revenge and retribution when others thwart our goals. Our faculty of sympathy, our ability to see the world from another's perspective, is an essential substrate of morality, but sympathy coexists with schadenfreude, pleasure that we feel in the pain of another, especially one who has wronged us.

Our moral instincts evolved because they promoted the differential survival and reproduction of our ancestors, relative to other individuals with other instincts, who survived less well and produced fewer offspring. For most of our ancestors, leaving genes to posterity depended on being accepted into social groups that maintained collective control over their own membership. Individuals who engaged in 'unacceptable' behaviors could be shunned, excluded from the enjoyment of public goods, expelled from the tribe, or even killed. Thus, the desire for acceptance and the dread of rejection are among the most powerful of human motivations.

Third-person sympathy is likely to have evolved because of the great benefits to be derived from within-group cooperation and the great cost of exclusion from those benefits. If reproduction were contingent on acceptance into a group, and acceptance was contingent on the expression of 'moral' behaviors and the suppression of 'immoral' behaviors, then humans may have socialized themselves by excluding carriers of genes for antisocial behaviors from their company.

If the prudent course is to seem to possess other-regarding preferences, then the most effective way to seem to possess such preferences may be to actually possess them. Prudence may have initially recommended the appearance of benevolence, but natural selection may then have favored genetic changes that conferred benevolence as a psychological motivation. The prudential reasons for benevolence could, in principle, be transferred entirely from the realm of utility and reason to the realm of fitness and instinct, leaving benevolence as a pure psychological motivation, untainted by self-interest. We could thereby look our fellows in the face and truly avow that we value their happiness as our own. In truth, self-interest remains a powerful psychological motivation, with self-love and benevolence co-existing in our psyches as instinctual motivations that are not fully reconciled.

The evolution of cooperation is often presented as the outcome of a tension between selection among groups and selection within groups (Sober and Wilson, 1998). In this view, competition among groups favors cooperation within groups, with the latter constantly undermined by a reproductive advantage for freeloaders who share the benefits of cooperation but do not pay the costs. The two levels of selection work in concert to enhance cooperation, however, if groups expel their 'antisocial' members or withhold the fruits of cooperation from those who do not contribute. Thus, the exigencies of between-group conflict may have been a powerful force for the enforcement of solidarity and moral policing within groups, if groups of

cooperative individuals were better able to expand and defend their territories, sometimes by the brutal extermination of their neighbors (Wrangham, 1999).

Natural selection may also have favored the psychological ability to negotiate truces, and avoid mutual destruction in escalated conflicts, but only if the inability to bury the hatchet was associated with reduced genetic reproduction at some level of between-group or within-group competition. Instinctive peace-makers may inherit the earth, but only if they leave more descendants than instinctive grudge-holders. Both parties benefit from a truce, but each party must be wary of shifts in the balance-of-power that favor opportunistic aggression by the other, and, a corollary less often acknowledged, must be ready to exploit opportunities for low-risk aggression when they arise.

4.2. Rational elements

> "But though reason is undoubtedly the source of the general rules of morality, and of all the moral judgments which we form by means of them; it is altogether absurd and unintelligible to suppose that the first perceptions of right and wrong can be derived from reason, even in those particular cases upon the existence of which the general rules are formed. . . . nothing can be agreeable or disagreeable for its own sake, which is not rendered such by immediate sense and feeling." (Smith, 1976, p. 320)

> "Those general rules of conduct, when they have been fixed in our mind by habitual reflection, are of great use in correcting the misrepresentations of self-love concerning what is fit and proper to be done in our particular situation." (Smith, 1976, p. 160)

Smith believed our moral intuitions come from our passions, but our moral rules come from reason. We derive general rules, not from the judgments of the man within our breast, a corruptible spectator who cannot always view our actions dispassionately and impartially,[10] but from the natural sentiments we feel when we observe the actions of others and observe others' judgments of these actions. In particular, we desire the admiration that comes from general approbation of our conduct and fear the opprobrium from general disapprobation. We derive our moral rules from the sentiments we experience in third-party observations of others, not from our simulated, and less-reliable, third-party observations of ourself (Smith, 1976, pp. 156–160).

Reason may be used to *rationalize* our behavior and our opinions; to justify positions, we have already adopted, to ourselves and to others. Haidt (2001) has argued that the principal use of reason in moral judgment is finding *post hoc* justifications for moral intuitions. In Haidt's view, moral reasoning has a limited power to change the behavior of others by re-framing questions and triggering new moral intuitions but is usually ineffective. But, reason can also be used to *reflect* on our reasons, to bring them into accord with each other, and to modify them in the light of experience. In Smith's view, self-reflective rules serve as a check on the unfettered expression of our moral intuitions. A commitment to behavior governed by moral rules helps us to avoid self-deception and acts as a brake on impulses we would later regret.

Reason may dictate fair behavior to be logically consistent. Sympathy has evolved the better to understand others, and reason the better to calculate costs and benefits of different options. In a situation in which I can benefit at your expense, I calculate my expected utility for each of my choices, and use sympathy to calculate your expected utility for each of your choices, the better for me to predict and counter your likely responses. The accuracy of my calculations depends on the quality of my simulation of your choices and the quality of my reasoning. Moreover, when I adopt a third-person perspective to anticipate the reactions of an observer, I elide the distinction between you and me. Why then, as a purely rational question, unconnected to emotion, should I value my utility higher than yours in choosing a course of action?

Emotion enters into my calculation of our respective utilities, and emotion may strongly advocate that my utility trumps your utility, but reason has considered the question from both sides and recognizes that the difference between *my* utility and *your* utility is an arbitrary criterion for breaking the fundamental symmetry of our situations.[11] How often people choose their behavior based on such abstract considerations is an open question, but rational arguments of this kind are frequently used in attempts to persuade others of what they ought to do.

4.3. Cultural elements

> "Since our sentiments concerning beauty of every kind, are so much influenced by custom and fashion, it cannot be expected, that those, concerning the beauty of conduct should be entirely exempted from the dominion of those principles." (Smith, 1976, p. 200)

[10] "When he is at hand, when he is present, the violence and injustice of our own selfish passions are sometimes sufficient to induce the man within the breast to make a report very different from what the real circumstances of the case are capable of authorising" (Smith, 1976, p. 159).

[11] For Smith, the special prominence given to our own interests was not arbitrary when viewed from the perspective of final causes: "Every man . . . is first and principally recommended to his own care; and every man is certainly, in every respect, fitter and abler to take care of himself than of any other person" (Smith 1976, p. 219). Man cares for himself because otherwise he would "have no motive for avoiding an accident which must necessarily diminish his utility both to himself and to society; and Nature, from her parental care of both, meant that he should anxiously avoid all such accidents" (Smith 1976, p. 148).

Our bodily sympathy and direct observations of others, our emulation of their behavior and sympathy with their thought, our attempts to persuade and our changes of mind, our desire for approval and fear of rejection, our listening to friends and to strangers, to parents and teachers, to rabbis, ayatollahs, and priests, and our reading of texts and viewing of film have caused the clay of our moral intuitions to be molded by interpersonal interactions and cultural processes. Moral thought and moral practice are influenced by untold generations of argument and reasoning about moral dilemmas.

Within social groups, the frequencies of alternative moral concepts wax and wane as a consequence of backsliding, persuasion, conversion, and execution of heretics. New concepts are proposed and old concepts mutate. These concepts become organized into coherent moral codes that differ in the sentiments they reward and the sentiments they punish, in the forms of sympathy they encourage and the forms they discourage. And the adherents of these codes, and those who live under a code's shadow, learn to modify their public selves accordingly. Conformity to explicit and implicit rules is a hallmark of most successful moral traditions, enforced by the fear of punishment and the fear of rejection. Among social groups, societies that were organized in ways that minimized internal conflicts and delivered more goods to their members possessed moral codes that were more likely to be emulated by other societies. And societies that were able to marshal military might to incorporate larger territories and larger populations under their moral code were more likely to have elements of their code copied by threatened neighbors.

A common feature of moral behavior is indignation against transgressors. Someone who is judged to have acted immorally has thereby placed himself outside the protection of the moral restraints we observe in our interactions with others. He deserves to be punished and made to feel pain. Many of the worst things done by humans to other humans have been perpetrated by groups who saw their actions as morally justified by the immoral behavior of their victims (Haig, 2007).

Moral codes are coercive. Their commands are considered universal and absolute by their adherents, and thus binding on all, regardless of personal preference. The 'culture wars' of the contemporary United States can be viewed as an escalated conflict between alternative moral codes. If a policy can be defined as moral, much energy can be mobilized to advance it. If it can be defined as immoral, the forces of moral outrage can be marshalled against it. For these reasons, political debate is often framed in moral terms, as arguments about what is right, especially by those who pursue selfish ends.

Moral codes are the products of cultural evolution and have evolved self-protective adaptations. Certain forms of sympathy are prohibited or discouraged, especially sympathy toward competing moral principles. For many, it is a sin to experience sympathetically, nay even to think about the possibility of experiencing, a moral principle from a different perspective. These prohibitions apply both to the 'politically correct' and the 'culturally conservative'. If I were to venture one suggestion as to how to move toward a 'truce' on the battlefield of competing moral 'absolutes' it would be to recognize, and then combat, our partisan prohibitions on sympathy.[12]

4.4. Responsibility

> "When I endeavour to examine my own conduct, when I endeavour to pass sentence upon it, and either to approve or condemn it, it is evident that, in all such cases, I divide myself, as it were, into two persons; and that I, the examiner and judge, represent a different character from that other I, the person whose character is examined into and judged of." (Smith, 1976, p. 113)

There are many voices, both external and internal, telling me what I should do. I am bombarded from without, by exhortations to do this or do that, by threats of punishment and promises of reward, by reasoned arguments and ecstatic visions, and I am beguiled from within, by reason, conscience, duty, honor, hopes and fears, contradictory passions, contradictory rules, and competing moral traditions that I have partially internalized. And, perhaps, there are silent voices of an unconscious self, working behind the scenes and responsible for my unaccountable impulses. Even different genes within my genome may come down on different sides of internal moral conflicts (Haig, 2003, 2006b). But when a decision is made "I" am responsible, as the arbiter among the stakeholders of the Self. May the Lord have mercy on my soul.[13]

Acknowledgements

Ideas are communally generated and this paper was no exception. It is a reformulation of things I have read and things I have heard from many people and to them, although they are not otherwise acknowledged, thanks. The paper has benefited from comments of Lucas Mix, Jonathan Wight, Elias Khalil, Eric Schliesser, James Simpson, three anonymous reviewers, and Adam Smith, whose insights have greatly improved the paper.

References

Alexander, R.D., 1987. The Biology of Moral Systems. Aldine de Gruyter, New York.
Arrow, K.J., 1963. Social Choice and Individual Values, second ed. Yale University Press, New Haven.

[12] Both sides fear that unreciprocated sympathy will be exploited. Politics is often trapped in a Prisoner's Dilemma in which moral intransigence is the dominant strategy.

[13] "And may the Lord have mercy on your soul," was spoken by English judges when passing sentence of death.

Baldwin, J.M., 1896. A new factor in evolution. American Naturalist 30, 441–451, 536–553.
Brown, V., 1994. Adam Smith's Discourse. Routledge, London.
Dawkins, R., 1982. The Extended Phenotype. Oxford University Press, Oxford.
Fogassi, L., 2011. The mirror neuron system: how cognitive functions emerge from motor organization. Journal of Economic Behavior and Organization 77, 66–75.
Gallese, V., 2007. Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. Philosophical Transactions of the Royal Society of London B 362, 659–669.
Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychological Review 108, 814–834.
Haig, D., 2003. On intrapersonal reciprocity. Evolution and Human Behavior 24, 418–425.
Haig, D., 2006a. The gene meme. In: Grafen, A., Ridley, M. (Eds.), Richard Dawkins. How a Scientist Changed the Way We Think. Oxford University Press, Oxford, pp. 50–65.
Haig, D., 2006b. Intrapersonal conflict. In: Jones, M.K., Fabian, A.C. (Eds.), Conflict. Cambridge University Press, Cambridge, pp. 8–22.
Haig, D., 2007. The amoral roots of morality. In: Steinberg, D. (Ed.), Biomedical Ethics. University Press of New England, Lebanon, NH, pp. 25–28.
Held, V., 2006. The Ethics of Care. Oxford University Press, Oxford.
Mayr, E., 1961. Cause and effect in biology. Science 134, 1501–1506.
Molnar-Szakacs, I., 2011. From actions to empathy and morality – A neural perspective. Journal of Economic Behavior and Organization 77, 76–85.
Nowak, M.A., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. Nature 393, 573–577.
Reddy, V., 2008. How Infants Know Minds. Harvard University Press, Cambridge, MA.
Richerson, P.J., Boyd, R., 2005. Not by Genes Alone. University of Chicago Press, Chicago.
Smith, A., 1976. The Theory of Moral Sentiments. Oxford University Press, Oxford.
Sober, E., Wilson, D.S., 1998. Unto Others. Harvard University Press, Cambridge, MA.
Tinbergen, N., 1963. On aims and methods of ethology. Zeitschrift für Tierpsychologie 20, 410–433.
Trivers, R., 2000. The elements of a scientific theory of self-deception. Annals of the New York Academy of Sciences 907, 114–131.
Wrangham, R.W., 1999. Evolution of coalitionary killing. Yearbook of Physical Anthropology 42, 1–30.