An Agent-Based Computational Approach to "The Adam Smith Problem"

Gavin, Michael

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Historical Social Research
Historische Sozialforschung

Michael Gavin:

An Agent-Based Computational Approach to "The Adam Smith Problem".

doi: 10.12759/hsr.43.2018.1.308-336

Published in:

Historical Social Research 43 (2018) 1

Cite as:

Gavin, Michael. 2018.
An Agent-Based Computational Approach to "The Adam Smith Problem".
Historical Social Research 43 (1): 308-36. doi: 10.12759/hsr.43.2018.1.308-336.

Historical Social Research
Historische Sozialforschung

All articles published in HSR Special Issue 43 (2018) 1:
Agent-Based Modeling in Social Science, History, and Philosophy.

An Agent-Based Computational Approach to "The Adam Smith Problem"

Michael Gavin *

Abstract: *»Das Adam Smith Problem. Ein agentenbasierter Integrationsvorschlag«.* This paper uses agent-based modeling to investigate the philosophy of Adam Smith. During his lifetime, Smith published two books. In An Inquiry into the Nature and Causes of the Wealth of Nations (1776), he argued that market behavior was dictated primarily by self-interest. However, his earlier book, The Theory of Moral Sentiments (1759), placed sympathy and benevolence at the center of human social psychology. Reconciling these apparently contradictory views is known as the Adam Smith Problem. This study uses an agent-based model to combine Smith's theories, exploring how social norms affect economic behavior, and vice versa. Although they share many basic assumptions, Smith's works leave important questions unanswered: Theory of Moral Sentiments offers a strong model of how social norms consolidate but lacks a coherent explanation for how norms change over time, and, on the other side, Wealth of Nations does not account for the influence of social norms on commercial transactions nor for the durability of seemingly irrational norms in a context of market competition. Considering both theories together in a single model sheds light on their underlying tensions and exposes instabilities that Smith did not anticipate.

Keywords: Adam Smith, agent-based simulation, agent-based computational economics, social norms, digital humanities, intellectual history, Theory of Moral Sentiments, Wealth of Nations.

1. Introduction

Scholars of eighteenth-century philosophy are likely, at one point or another, to bump up against the "Adam Smith Problem." The contours of this problem are familiar but worth reviewing. During his lifetime, Adam Smith published two books: *The Theory of Moral Sentiments* (1759) and *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776). In *Wealth of Nations* he argued that market behavior was dictated primarily by self-interest. Smith writes,

* Michael Gavin, Department of English, Welsh Humanities Building, University of South Carolina, Columbia, SC 29206 USA; mgavin@mailbox.sc.edu.

Historical Social Research 43 (2018) 1, 308-336 | published by GESIS
DOI: 10.12759/hsr.43.2018.1.308-336

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages (I.2.2.).

However, Smith's other major work, *The Theory of Moral Sentiments* (1759), begins by offering a fundamentally different view of human motivation:

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it (I.i.1.).

By sympathetically identifying with others, Smith argues, people develop an internal sense of justice, propriety, and benevolence that informs all their actions. While in *Wealth of Nations* Smith makes little mention of sympathy, in *Theory of Moral Sentiments* he places sympathy at the center of human social psychology. Moral behavior and market behavior are presented as separate phenomena. Smith makes little effort in either book to reconcile their apparent contradictions, nor does he ever explicitly justify his decision to treat his topics separately.

How do Smith's moral sentiments emerge into collective norms, and how do those norms affect commercial behavior? Defenders of Smith have argued that his theories share a common abstract premise: both take the individual person – the feeling, desiring, and choosing agent – as their basic unit, and both are concerned primarily with how interactions among agents cause larger social patterns. However, whereas other scholars are content to acquit (or accuse) Smith based on the fundamental compatibility (or incompatibility) between his assumptions, this essay tackles the much trickier problem of evaluating whether those behavioral assumptions entail compatible social consequences. I use an agent-based model for this purpose. My argument proceeds with a brief discussion of agent-based modeling as a technique for intellectual history. I next review scholarship on Smith's philosophy. I then turn to the model itself, comparing its algorithms to Smith's theories and describing ways in which its results support or undermine various interpretations that Smith scholars have put forward. I will argue that, although they share many basic assumptions, Smith's works leave important questions unanswered: that *Theory of Moral Sentiments* offers a strong model of how social norms consolidate but lacks a coherent explanation for how norms change over time, and, on the other side, that *Wealth of Nations* does not account for the influence of social norms on commercial transactions nor for the durability of seemingly irrational norms in a context of market competition. Considering both theories together in a single model sheds light on their underlying tensions and exposes instabilities that Smith did not anticipate.

2. Agent-Based Computation and the Study of Social Norms

"Moral Markets" was designed and written in NetLogo (version 5.0.5).[1] Like similar toolkits, NetLogo allows researchers to simulate social and biological systems as fields of interacting entities (agents) whose simple individual behaviors collectively generate larger emergent patterns.[2] Joshua Epstein, a leading advocate and practitioner of agent-based modeling, refers to the technique as "generative social science."[3] Whereas traditional empirical research tests theories by comparing their predicted outcomes to real-world experiments or events, agent-based simulations test theories by using them as design principles for computer programs, then executing those programs and evaluating the results. If the effects we observe really do depend on the causes we posit, we ought to be able to replicate them *in silico*. This theory of explanation has been summed up by Fred Dretske in the motto, "If you can't make one, you don't know how it works" (Dretske 1994, 468).[4]

Applications of agent-based modeling to intellectual history remain rare, but a computer model of Smith's *Theory of Moral Sentiments* is not without precedent. The computational modeling of norms is fairly well-established in the social sciences, where Robert Axelrod has been a pioneering figure. His essay, "An Evolutionary Approach to Norms," uses game theory to model acts of moral defiance and punishment in an abstract, simulated community (Axelrod 1997). Following Axelrod, Epstein has designed a number of agent-based simulations to study norms: his models have outlined norms' relation to class distinctions, demonstrated how "cooperative persistence" might diffuse over space and time, and identified an inverse relationship between mental deliberation and conformity (Epstein 2006, 218). More recently, the 2014 collection of essays, *Minding Norms*, uses an agent-oriented approach to explore the psychology of conformity to show how "emergent" patterns of norm-formation correlate to "immergent" effects on individuals, who internalize moral beliefs through social interaction (Conte, Andrighetto and Campennì 2014, 9). In fact,

1 For an overview of NetLogo and its use as a tool for scientific inquiry, see Railsback and Grimm (2012) and Gilbert (2008).

2 A general overview of the theory and practice of agent-based modeling and its application in the social sciences can be found in Miller and Page (2007).

3 Most pertinently, see Epstein (2006). This collection of essays builds on Epstein's earlier work, described in Epstein and Axtell (1996).

4 Epstein carries this line of argument even further, suggesting that how-to-build-it knowledge is sufficient – not just necessary – for how-it-works knowledge. In *Growing Artificial Societies*, Epstein, with Axtell, asks, "What constitutes an explanation of an observed social phenomenon? Perhaps one day people will interpret the question, 'Can you explain it?' as asking 'Can you grow it?'" (20).

part of the *Minding Norms* project involved the creation of HUME1.0, an agent-based simulation premised loosely on the moral philosophy of Smith's predecessor, David Hume (Will and Hegselsmann 2014).

Modeling social norms doesn't involve empirical inquiry in the colloquial sense, in part because sociological data about moral beliefs is notoriously hard to collect, but more fundamentally because simulation always remains at an ontological arm's-length from reality.[5] What happens in a simulation never happens outside the simulation. For scientists, this gap can be very troubling, and much polemical effort has been invested in critiquing and defending the use of models and simulations. (The science of climate change, for example, depends crucially on simulation.) Agent-based models of social systems have, thankfully, avoided being caught up in these controversies, and so its theorists are more free to acknowledge ABMs function as a "theory building" or heuristic tool. Rather than represent complex systems mimetically, agent-based simulations always point centripetally to the ideas at their core.[6] Models are especially useful when scholars have identified some simple process and hope to tease out its more complex secondary and tertiary consequences. For this reason, although the output of simulations involves statistical analysis, charts, graphs, and the other trappings of inductive reasoning, in fact models of this kind operate more like thought experiments extended through computation. They typically function deductively or abductively, and their purpose is to test the internal consistency of causal explanations (Bonnet 2007; Dixon 2012).

I began by creating a simulation of Smith's *Theory of Moral Sentiments*, then I combined that model with an already existing (indeed, famous and oft-studied) model in the field of agent-based computational economics. Based on my reading of the primary and secondary literature, to which I will turn in a moment, my guiding assumption was that Smith's philosophy is neither inherently flawed nor contradictory, but that the connection between his works is underdetermined. Therefore, the model was designed to exhibit behaviors at both the micro and macro levels that correspond analogically to Smith's ideas while "generating further, more specific, 'middle range' theories" about how his individual- and community-level generalizations correlate (Gilbert 2008,

[5] For a discussion of the ontology and epistemology of simulation in relation to the practice of history, see Gavin (2014). For a general account of simulation as a tool of explanation in the sciences, see Winsberg (2010).

[6] Jim Doran and Nigel Gilbert (1994) write, "There are different reasons for modelling. The aim may merely be to construct a model that is valid at least to some degree – that is, whose behaviour does match that of the target in at least some significant respects. By constructing such a model, we may hope to gain insights into the target itself. Such modelling is exploratory in nature, often involves theory building, and is the type of modelling mainly at issue in this book." Adam Smith's theory of moral sentiments tends to operate at a very abstract level, and so an agent-based simulation that uses Smith's theories as design principles must be similarly abstract.

41). That is to say, I hope to make better and more detailed inferences about how Smith's ethics inform his economics, and vice versa. The model is not meant to be validated against any dataset; instead its success depends on this heuristic result.

3. The Adam Smith Problem

3.1 Competing Interpretations of Smith's Work

Scholarship on the Adam Smith Problem can be divided into two camps. *Compatibilist* arguments identify common ground across the two works, while *incompatibilist* arguments look for points where the works would benefit from more explicit differentiation. Scholars tend to emphasize one view over the other, but most adopt a mix of compatibilist and incompatibilist positions. As originally formulated in the nineteenth-century, the Adam Smith Problem was understood straightforwardly as a contradiction between sympathetic altruism and market-based egoism.[7] As we shall see, this apparent contradiction largely disappears upon closer inspection, but tensions between Smith's works remain a topic of debate. According to Jack Russell Weinstein,

> In the last decades, […] [scholars] have replaced the original formulation with a host of other Adam Smith Problems, a diverse list of supposed incompatibilities between *TMS* and *WN* that call the unity of Smith's work into question (2013, 50).

(In what follows, my conclusions will add to this list.)

The strongest arguments for Smith's unity draw out the implicit conceptual similarities between his works.[8] Contra incompatibilist claims that altruism and egoism are opposites, Smith advanced an individual-based model of sympathy and moral development. The "sympathy" that individuals feel drives their socialization: by witnessing how other people handle joys and pains, we constantly evaluate them, and in doing so we internalize a sense of how others will judge us. Smith personifies this internalized socialization as an "impartial spectator":

> Whatever judgment we can form concerning [our own sentiments and motives], accordingly, must always bear some secret reference, either to what are, or to what, upon a certain condition, would be, or to what, we imagine, ought

[7] For an overview of the history of the Adam Smith Problem, see Otteson (2002).

[8] Another important compatibilist argument is biographical. Although they were first published almost twenty years apart (leading some scholars to believe that Smith had simply changed his mind), Smith revised and republished *Theory of Moral Sentiments* throughout the rest of his life and, as A. L. Macfie notes, "there is no hint that Smith himself thought there was any conflict of doctrine between the two books" (1967, 76).

> to be the judgment of others. We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. If, upon placing ourselves in his situation, we thoroughly enter into all the passions and motives which influenced it, we approve of it, by sympathy with the approbation of this supposed equitable judge. If otherwise, we enter into his disapprobation, and condemn it. (III.i.2)

Thus, what folk psychology might call a person's "conscience" becomes, for Smith, an internally projected personification of social norms.[9] By imagining ourselves under "the judgment of others" we learn to regulate our conduct and, Smith believes, develop a sense of common morality:

> Man naturally desires, not only to be loved, but to be lovely; or to be that thing which is the natural and proper object of love. He naturally dreads, not only to be hated, but to be hateful; or to be that thing which is the natural and proper object of hatred. (III.i.8)

Because people naturally seek to avoid admonishment – not only from other people but from their internally held ideas about what other people might say – they are guided by external and internal stimuli to obey "natural" virtues like justice, prudence, charity, and magnanimity.

3.2 Smith's View of Norms as Emergent Phenomena

Working out from these psychological assumptions, Smith argues that moral standards develop much like commercial markets. Widely shared social norms represent collective patterns that emerge through countless personal encounters, as people trade judgments about proper and improper behavior. James Otteson places this market analogy at the center of his claim that Smith's theory describes a world "in which free exchanges among participating people give rise, over time, to an unintended system of order" (2002, 101).[10] Genevieve Lloyd describes the impartial spectator's emergent properties in similar terms: "out of those shifting and turbulent – intense yet fragile – interconnections between selves, there can emerge a stable, objective moral consciousness" (2013, 92).[11] Across the two books, then, it's possible to trace Smith's persis-

---

[9] Knud Haakonssen describes the process this way: "The internal spectator has the force to prompt such adjustment of behavior as would otherwise be demanded by external spectators in order to satisfy the inclination to or the need for agreement or conformity. In other words, one only learns to see oneself as a person and as a member of a moral universe of agents through sympathy with others' view of one's identity and situation in the world. Society is, as Smith says, the mirror in which one catches sight of oneself, morally speaking" (2002, xv).

[10] Weinstein disagrees with Otteson's "market metaphor." Weinstein points out that "Smith's moral theory does not involve *exchange*" (*Adam Smith's Pluralism*, 66, original emphasis).

[11] For Lloyd, the central puzzle of Smith's ethics is the question of how such a system can give rise to a shared sense of moral standards without collapsing into moral relativism. Amartya Sen (2009, 403-6) distinguishes between two kinds of impartiality: "closed impartiality,"

tent concern with patterns of self-organization. Both *Wealth of Nations* and *Theory of Moral Sentiments* are important founding documents in the study of complex social systems, and they share a common underlying thesis that individuals and societies are mutually constitutive.

Much like markets reach equilibrium, morals reach consensus, although this consensus is neither perfectly stable nor absolute. Jerry Evensky explains, "In Smith's analysis, every society is a unique social construction with its own established norms that are peculiar to its time, place, and circumstance" (2005, 49). Smith does not endorse moral relativism, but his theory is acutely aware of cultural difference. In a chapter "Of the Influence of Custom and Fashion on Moral Sentiments," Smith writes:

> Every age and country look upon that degree of each quality, which is commonly to be met with in those who are esteemed among themselves, as the golden mean of that particular talent or virtue. And as this varies, according as their different circumstances render different qualities more or less habitual to them, their sentiments concerning the exact propriety of character and behaviour vary accordingly. (V.I.17)

Like the "invisible hand" that guides an economy's production and circulation of goods, the "golden mean" of moral propriety is subject to continual flux while, as Smith makes clear elsewhere, tending always toward a common good. The same spirit that motivates Smith's optimistic liberalism in *Wealth of Nations* leads him in *Theory of Moral Sentiments* to a bottom-up system of ethics that remains open and pluralistic while affirming "natural" virtues that transcend cultural boundaries.

For scholars who read *Theory of Moral Sentiments* closely, the Adam Smith Problem can seem to melt away. D.D. Raphael and A.L. Macfie called it a "pseudo-problem based on ignorance and misunderstanding" (Smith 1759 [1982], 20). More recently, important books by Amartya Sen, Charles Griswold, James Otteson, Jerry Evensky, Stephen McKenna, and Jack Russell Weinstein have emphasized strong continuities across Smith's corpus, even if they sometimes disagree whether to call the "tensions" and "inconsistencies" that remain a "problem."[12] Griswold (1999) finds a conflict between Smith's

---

which evaluates questions of justice according to local, parochial social norms, and "open impartiality," which self-consciously adopts a more cosmopolitan outlook. Sen argues that Smith advocates open, cosmopolitan impartiality.

[12] The word "problem" in the phrase "Adam Smith Problem" perturbs some Smith scholars. Although he disputes the phrase "Adam Smith Problem," Macfie acknowledges "considerable inconsistencies" between the two books (1967, 81). Similarly, Weinstein (2013) dismisses the notion of a "Problem" with bravado, saying that its resolution is "shockingly straightforward." Yet, he too acknowledges "there are tensions, sure." It's rarely clear what's at stake in the debate over the phrase. Over the past thirty years, the dispute seems to hinge on whether Smith's works share a fundamental unity, despite many differences, or whether Smith's works have many differences, despite sharing a fundamental unity.

"historicism" and his "Platonic commitments" to moral truths. Göçmen argues that Smith's works expose a fundamental flaw in the logic of capitalism, and that *Theory of Moral Sentiments* offers an "immanent critique" of commercial society (2007, 157).[13] Most pertinently, Otteson (2002) points out that Smith fails, at a very basic level, to explain "the matter of how morality mixes with markets." "If the Smith of TMS is right," Otteson says, "one does not – or should not – check one's morality at the marketplace door."

In the Moral Markets model I follow Otteson's call to account for the way "morality mixes with markets" in Smith's philosophy. If the compatibilist view of Smith's work is correct, *Theory of Moral Sentiments* and *Wealth of Nations* share a common model of human behavior. In which case, it should be possible to formalize Smith's moral and economic behavioral theories in a way that demonstrates their fundamental compatibility while leaving open the question of whether he was right about their emergent social properties. My goal is to paraphrase Smith's views into machine-readable algorithms that test how moralizing behaviors influence market behaviors *in silico*. If Smith's models of human behavior share many similarities, will they behave well together in simulation?

3.3 Smith's Idea of 'Natural Prices' Compromised by Moral Judgments

Before turning to the model, however, a few more words about the Adam Smith Problem. Otteson's answer to the question of how morality mixes with markets rests on what he calls "the familiarity principle." Because Smith's benevolence is generally confined to persons within an individual's circle of family and friends, and because commercial transactions happen largely among strangers, Otteson believes that self-interest, rather than sympathy, can account for economic transactions. Macfie and Raphael (1982) and Weinstein (2013) make similar arguments. However, for Smith, moral sentiments are infused in all social interactions. The impartial spectator is deeply embedded in the Smithian psyche, and he provides no hint that it's ever possible to set one's socialization aside.

Consider, for example, Smith's brief discussion in *Wealth of Nations* of wage and profit differentials in eighteenth-century Britain. In a chapter titled "Of Wages and Profit in the Different Employments of Labour and Stock," he

[13] Göçmen explains, "My main claim is that there is only one concept of human nature in Smith's work, but that it consists of two complementary elements. The first is a general normative view of human nature. The second and more specific is an account of the human situation in commercial society. There is indeed a contradiction between these two aspects of Smith's anthropological view. Unlike many scholars, however, I suggest that this contradiction should not be ascribed conceptually to Smith. Rather, it is a real problem arising from social relations in commercial society" (2007, 2).

includes a section that explains "Inequalities arising from the Nature of the Employments themselves." His thoughts lead him to contrast "honourable professions" from trades that involve cutting meat:

> Honour makes a great part of the reward of all honourable professions. In point of pecuniary gain, all things considered, they are generally under-recompensed, as I shall endeavour to show by and by. Disgrace has the contrary effect. The trade of a butcher is a brutal and an odious business; but it is in most places more profitable than the greater part of common trades. The most detestable of all employments, that of public executioner, is, in proportion to the quantity of work done, better paid than any common trade whatever. (I.10.5)

Self-interest guides this economic account, but that self-interest involves a complicated amalgam of motives: Smith counts "honour" and "disgrace" in the wages of labor. These concepts are moral sentiments, and their circulation has a lot to do with how labor and goods are priced in a commercial society. When Smith says that public executioners earn higher wages "in proportion to the quantity of work done" because their work is "detestable," he endorses an economic model that factors emotions and morals into the components of price.[14]

For now I just want to highlight one point: scholars puzzling over the Adam Smith Problem tend to focus exclusively on the narrow issue of "benevolence," and in doing so they radically underestimate the applicability of Smith's ethics to his economics. We may not be able to rely solely on the benevolence of our butcher for our dinner, but he'll feel every moment the sting of our disdain, and his work is priced accordingly. The executioner demands a special price for human heads. Smith provides other examples. Innkeepers make a higher rate of profit because they endure the rude behavior of guests. Clergymen make less because they enjoy the dignity of their office. Opera singers and actresses earn exorbitant wages because they must overcome the embarrassment of public performance, which the accomplished ladies of Smith's eighteenth-century Britain would sooner forgo. These examples are not exhaustive but are meant to demonstrate Smith's general point: all wages and profits are naturally determined at least in part by the moral sentiments that attend their employment of labor and stock.

---

[14] The account Smith offers here shares much in common with "rational-choice theory," which was first popularized by economist Gary Becker in *The Economics of Discrimination* (1957). Trying to account for the way that bigotry depresses wages for black workers, Becker factored in a "taste for discrimination" that racist whites sometimes hold, and he developed a model to describe how such "disgrace" (to return to Smith's phrase) is factored into the cost of labor. The core idea of rational-choice theory is that putatively non-rational, emotional, or cultural considerations can be usefully reduced to factors in a straightforward decision-making process about cost and benefit, or risk and reward.

However, such remarks appear only briefly in *Wealth of Nations*, and Smith never really explains how he believes moral ideas affect commercial exchange. Indeed, his definition of the "natural price" of goods largely sets the matter aside:

> When the price of any commodity is neither more nor less than what is sufficient to pay the rent of the land, the wages of the labour, and the profits of the stock employed in raising, preparing, and bringing it to market, according to their natural rates, the commodity is then sold for what may be called its natural price. (I.7.4)

Price covers all the major costs: rent of the land, wages of labor, and the profits of stock. The natural price of any commodity is achieved when its market approaches equilibrium according to their natural rates. At first glance, this definition looks like a simple tautology. Smith makes no mention here of the "natural inequalities" of wages and profit that attend different employments, but neither does he explicitly exclude them, so it could be that they are implicitly included in the phrase "according to their natural rates." If so, the moral sentiments that accompany labor and profit would somehow be embedded in the prices of commodities and therefore of all goods and services, meaning that any commodity's "natural price" is set ultimately by the "natural virtues" described in *Theory of Moral Sentiments*. But this presents a difficult problem. On the one hand, Smith suggests that exchange value is highly dependent on moral value. On the other hand, he provides no mechanism for identifying either the degree of dependence or the micro-level process. How are prices affected? How do people make choices? The dots are tantalizingly close together, but Smith never connects them. One way to understand the Adam Smith Problem is to see it as the scholarly project of connecting these dots.

Because Smith left so many of the connections between his works implied, readers interested in those connections are forced to make inferences on Smith's behalf. The question shifts from "What did Smith argue?" to "What do Smith's ideas *imply*?" Generative modeling is particularly well-suited to contribute to this line of inquiry, because it was specifically designed to facilitate this kind of inference.

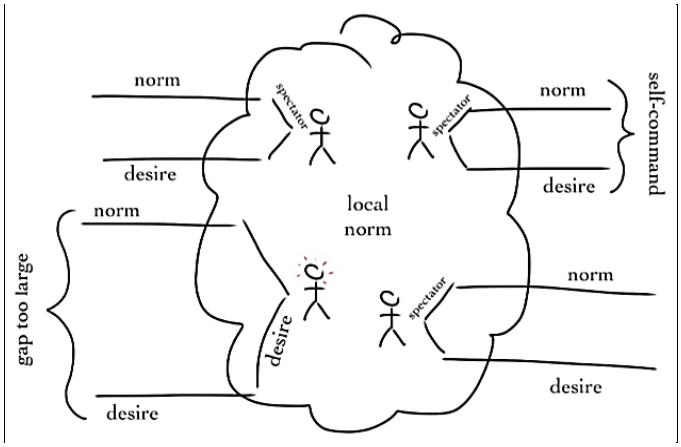
4. Formalizing Smith's Ethics

4.1 Review of Smith's 'Impartial Spectator' Model

According to Smith's *Theory*, every person is pulled simultaneously by three distinct moral forces. First, they have internally felt pleasures and pains that motivate them to act on passions, some of which are "selfish" and some "sociable." Second, they have the examples and admonishments of their immediate interlocutors, which together amount to locally determined norms of behav-

ior. Lastly, every person carries within their minds an impartial spectator who speaks on behalf of their more durably experienced moral beliefs: the impartial spectator is the "golden mean," or average of averages, of every local norm that people encounter over the courses of their lives.

Figure 1: The Impartial Spectator Procedure



The impartial spectator procedure is described in Figure 1. It is triggered whenever individuals encounter one another. Because all social encounters are occasions of moral surveillance, people feel pressure to exert magnanimity (or "self-command") at every moment. These pressures may involve significant ethical dilemmas, such as deciding whether to cheat a customer, but they might just as well involve simpler acts of self-control, like suppressing immoderate laughter or using a handkerchief, rather than a petticoat, to wipe your nose. The pressure to conform will prove more or less powerful depending on the extent to which a person's desires deviate from social demands. For most people most of the time, Smith suggests, conformity isn't particularly difficult (and may even escape notice). The gap between one's desires and one's context usually is not too severe, and so people behave most of the time according to the dictates of their impartial spectators without paying it much heed. Further, because people exist together in a common social space – because they share a common "age and nation"– their impartial spectators tend toward a common mean and emerge as natural standards of behavior that are reinforced synchronically, by cross-cultural comparison, and diachronically, through education. However, neither the standards nor their enforcements are absolute, so individuals with unusual desires or poor self-command will always pop up here and there, and they may at times even disrupt the entire system, which is subject to an unspecified process of change over time.

4.2 Parameters of the Moral Markets Model

The above summary of Smith's theory is neither original nor, I hope, controversial. However, my next claim pushes against the general drift of recent scholarship on Smith's philosophy.[15] The ethical subject described in *Theory of Moral Sentiments* is mathematical and procedural at its base, and therefore it translates easily into algorithmic expression. The impartial spectator procedure describes the interaction of four closely related, commensurable variables:

- Agents' desires (*D*), assigned randomly to signify the agent's desired behaviors at any given moment;
- Agents' self-control (*M*, for "magnanimity"), a randomly assigned value meant to signify that agent's ability to conform in any given situation;
- Agents' behavior (*B*), their actual behavior at any given moment;
- Agents' moral belief (*S*, for "spectator"), their ideas about how agents ought to behave in general.

In Smith's theory, people are always comparing themselves to those around them. At any given moment, they experience a local norm that prevails among their neighbors, represented in the model as the average behavior of people within a visible area (N_a). The impartial spectator for each person is the "golden mean" of the local norms they individually experience over time.

Once these variables are in place, they can form the core of a simple algorithm that models behavior. At every time step, the model updates its key variables. The agents move randomly (Smith offers no theory of space or movement, so that's left completely abstract). Every area will host a new norm (N_a) based on the average behavior of agents within a visible radius

$$N_a = \text{mean}(B_a)$$

Every agent updates its moral belief by incorporating the local norm (N_l) into a rolling average of its personal experience, such that

$$S_i = \text{mean}(N_{ai})$$

After these values are determined, agents respond to their surroundings by adjusting their outward behavior to conform (or not) with agents around them. People sometimes find it difficult to conform to their internally held sense of propriety; this difficulty can be expressed as the absolute difference between the demands of local propriety and their internally felt desire ($| N - D |$). If the gap between social expectations and internal passions is less than or equal to their magnanimity (*M*), they will behave according to their impartial spectators; otherwise they will follow their socially unacceptable desires. Expressed algorithmically:

[15] In addition to Weinstein, see Evensky (2005, 247): "Smith sees humankind as a uniquely complex realm of nature that does not lend itself to such reductionism."

For every person p at time t in area a,

$$\text{if } | N - D | <= M$$
$$\text{then } B = S$$
$$\text{otherwise } B = D$$

Formalizing the impartial spectator procedure in this way allows for a number of computationally verifiable inferences. Because the outcome of each person's decision about how to behave, B, contributes to the social norms, N, that locally prevail, this algorithm operates recursively ("poietically," in Charles Griswold's phrase) to generate its own conditions.[16]

4.3 The Impartial Spectator in Action: Testing the Model

In *Theory of Moral Sentiments*, Smith does not speculate about precise sociological facts. It's never clear what percentage of the population (a phrase anachronistically applied to Smith's thinking) he believes conforms to natural virtues. However, the general mechanism is clearly stated: moral sentiments emerge through their expression and regress toward a mean as they are variously internalized and assented to. All else being equal for a set of randomly interacting agents, this algorithm should tend inexorably toward values of B, N, and S with standard deviations (σB, σN, and σS) that correlate directly with the average values of M and a. Restated in common prose, this means that if people in a community have greater self-control and less privacy, their communities will have a more uniform moral composition. If individuals prefer not to conform or cluster into small subgroups, their communities will display a greater diversity of behavior and moral beliefs.[17] If we accept the above summary of

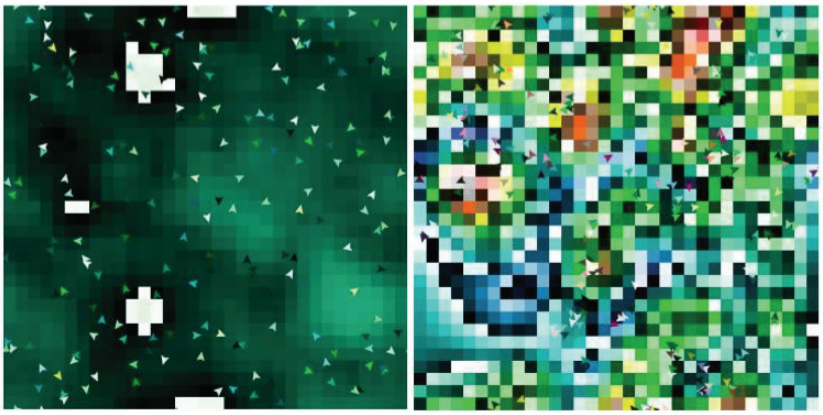
[16] Griswold (1999, 146) writes, "From Smith's spectating standpoint in the philosophical critic's balcony of the *theatrum mundi*, the (so to speak) non-natural nature of moral standards is inseparable from the fact that all of morality, and indeed all of the human 'world,' is a complex whole that we communally impose on ourselves. Yet Smith also holds that this 'poietic' character of the 'world' is not generally visible either to actors or spectators. As a result, morals are ordinarily thought to possess a reality and authority that is external to us." The recursive element of norm formation and internalization is also central to the agent-based computational model described in Conte, Andrighetto, and Campenni (2014, 8-9): "While moving around in the social space, meeting with one another, [agents] act as intelligent norm-seekers, or norm-detectives: they apply some of their cognitive capacities to finding out norms. Once external signals have been turned into mental representations of a normative kind, agents may decide whether to behave in accordance with norms, or with what they believe to be such, or not. By doing so, they provide new inputs to other agents, who will possibly convert them to the corresponding behaviors, and so on and so forth, in a recursive way."

[17] The contrast between cosmopolitan and parochial patterns of norm emergence, implied in Smith's theory, is noted in Sen (2009). Göçmen (2007, 134) finds similar implications: "A harmonious and open society without fragmentation can form our emotional and intellectual dispositions in a different way from a closed one. Under the conditions of an open and harmonious society, for example, the general intellectual and bodily capacities of man can

Smith's views, and if the consequences that logically seem to follow from that summary bear out *in silico*, we have good reason to believe that Smith's ideas imply similar consequences.

Moral Markets began as a simple model of Smith's ethics, based on the impartial spectator rule described above. To ease visualization, I indexed the agents' variables (*B*, *D*, and *S*) to a scale of 0 to 140, which happens to match NetLogo's color scheme. Thus, as agents change their moral self-presentation (*B*), they change color. The patches represent local norms (*N*) and they change color to match the colors of agents in their area, so the shifting normative standards can be visualized as changing colors in a field. When the model runs, its colors ebb and flow like weather-map clouds looping over a blank landscape or like the bubbling oils of a lava lamp. Through the graphical interface, users can test how the system responds to changes in agents' parameters (see Figure 2.) When agents are initialized with high levels of magnanimity, they tend to conform to each other's expectations, and the view takes on a fairly even color with only slight variations. When agents feel free to follow their own passions, the system erupts into particolored clusters.

Figure 2: Moral Markets



Moral markets (2014 NetLogo Community Model), exported Views after 300 Ticks Left: Self-command (M) and visibility (a) set to 100 and 8. With high visibility and high self-control, agents tend to conform. Right: Self-command (M) and visibility (a) set to 30 and 2. At lower levels of visibility and control, behavior is more diverse.

be formed in a way such that the agent can act spontaneously, i.e. without any long and deep considerations, on the basis of the principle of impartiality, spontaneously taking both general and particular interests into account. In a fragmented and, therefore, closed society in which the agent is embedded, his durable intellectual and emotional dispositions will be formed partially, so that the agent can act only from a partial point of view, being unable to take into account the general point of view."

Smith never speculates about how powerful the impartial spectator generally is, suggesting only that it varies from person to person. For this reason, in Moral Markets, each agent carries a different value of *M*, but those values are distributed along a normal curve that the researcher controls. It's possible to set the average *M* anywhere from 0 (meaning very few agents would ever conform to their moral beliefs) to 140 (resulting in total conformity). Testing the performance of a model across a spectrum like this is called "sensitivity analysis." Researchers test agent-based models by running them with different settings to see how sensitive their performance is to changing parameters. Sensitivity analysis allows researchers to identify points where models "break" (that is, where outcomes are completely unrealistic), as well as tipping points where "phase changes" in plausible emergent patterns occur.[18]

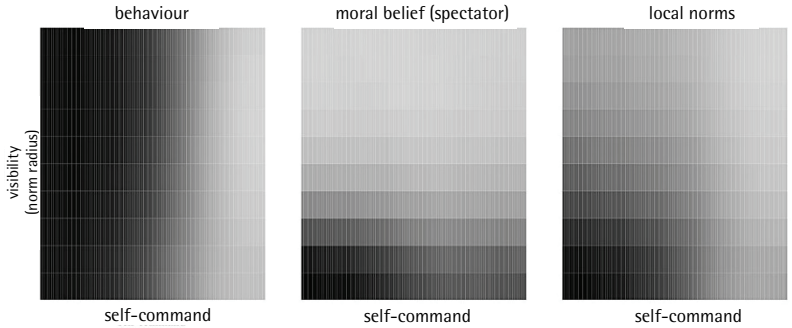
Visualizations like Figure 2 are helpful for getting an initial sense of how a system operates, and they're indispensable during the process of building a model because of how quickly they expose semantic errors in the code. However, they're of limited interpretive value if you want to make detailed comparisons across multiple runs. For that, it's better to sift through the data generated by the simulation. In this model, differences in how norms circulate correspond to differences in several key statistics (see Table 1). Color variations like those visualized in Figure 2 can be measured simply as standard deviations: How much variance is there in the agents' colors (their individual behaviors, *B*), in the colors of the patches (the local norm, or the average behavior in any given locale, *N*), and in each agent's invisible moral beliefs (the average of all norms experienced by agent, their "impartial spectator," *S*)? When the average self-control, or "magnanimity" is set to 100 (out of 140 total possible), most agents will conform to the dictates of their conscience in all but the most extreme scenarios. Indeed, whether visibility was high or low, conformity was at or near 100%. When their self-control was set only to 30, on the other hand, agents were much less likely to conform, and the behaviors they displayed were subject to much higher variation. By contrast, their moral beliefs (the "impartial spectator") varied by the agents' visibility. When the visible area is set only to two patches, rather than eight, agents measure their behavior against much smaller subsets of agents. Smaller groups allow for more variation, both in local norms and moral beliefs.

One striking fact about agent and patch behavior in "Moral Markets" is the persistent gap between levels of variation in behaviors (σB), local norms (σN), and sentiments (σS). This dynamic can be seen more clearly in Figure 3, which performs a more complex sensitivity analysis. Table 1 displayed the results of 4,000 runs of the model, 1,000 each at four different settings. Figure 3 shows

[18] According to Gilbert (2008, 45), sensitivity analysis is "aimed at understanding the conditions under which the model yields the expected results."

results of a 10,000 runs. Values of M were tested at every integer from 1 to 100, and, at each step, the model was run with values of a set from 1 to 10, in increments of 0.1. In the heat maps, M is distributed along the x-axis and a reaches up the y-axis. Dark colors on the maps show high levels of variance, and light colors represent low variance.

Figure 3: Self-Command and Visibility



Sensitivity analysis testing relation between self-command (M, x-axis) and visibility (a, y-axis). 3.1 shows that variation in behavior B is related almost entirely to self-command, and 3.2 show that variation in internal beliefs of the spectator (S) depend mostly on the visibility of agents to each other. The locally prevailing social norms (N) are sensitive to both variables.

As you can see in Figure 3.1, the left side of the map is entirely red [/dark]. At low levels of magnanimity, agents will display a wide range of behavior: few people bother to follow their own rules. At higher levels of magnanimity, they display low variation. This pattern holds over the entire vertical axis. In Figure 3.2 we see the opposite pattern. Variations in moral beliefs are fairly consistent horizontally but differ based on the value of a. Moral sentiments consolidate independently of actual behavior, but they display more or less variation depending on how visible agents are to each other. When visibility is set to low levels, agents learn parochially, comparing themselves only to other agents within a small radius. When visibility is set higher, agents develop in a more cosmopolitan way, taking large surveys of their peers at every step. How agents behave is highly dependent on their self-control, but what they believe isn't. Social norms, however, are responsive to both variables. In Figure 3.3, only the lower left corner displays a high level of variation. The areas on the upper left and lower right, which are darker for the values *B* and *S*, here appear light, suggesting conditions under which collective behaviors might consolidate into norms, even in the absence of widespread conformity, or even in a context of moral disagreement.

4.4 Initial Results: Using the Model to Reconsider Smith's Idea of 'Natural' Virtues

Still setting aside the incorporation of market behavior, to which I'll turn in the next section, these results point to a much-discussed interpretive difficulty in Smith's *Theory of Moral Sentiments*. This difficulty involves the relationship between morals and norms: To what extent does Smith's theory of "natural" virtues correspond to a theory of moral absolutes, and to what extent does his cultural pluralism lend support to "moral relativism"? Related to this is a separate but equally pressing question: Does Smith offer a merely descriptive account of how moral sentiments consolidate into social norms, or does he provide a trustworthy prescriptive ethics to evaluate moral dilemmas? An agent-based model can't answer these questions, but it can shed a different light on their central ambiguity, and it might suggest why Smith is so difficult to pin down.

In a loosely conforming society, this model suggests, social norms exhibit greater variety than morals. In a highly conforming society, less. In either case, individually held beliefs tend to have greater stability and exhibit less volatility over time. If we can imagine what the agents might say were they sentient, it's easy to imagine them reporting that morals are more stable, more true, than social norms – indeed, that morals transcend mere norms, which either vary from one context to the next or, at another extreme, artificially constrain everybody. Under Smith's theory, this suggests, people can believe the truth of moral sentiments and reach a widely shared consensus about them, even if few people actually abide by them. This analysis doesn't suggest that justice, benevolence, magnanimity, and prudence are "natural" virtues, but it does suggest why a practical thinker like Smith would be comfortable with the ambiguity involved in calling them so.

Some scholars have questioned whether Smith's theory of the impartial spectator is sufficient to account for the consolidation of a shared normative consensus.[19] To the extent that this agent-based simulation accurately executes Smith's assumptions, it supports his theory. Not only is the impartial spectator sufficient to generate, *in silico*, a widely stable normative field, the simulation also paints a very interesting picture about how those norms relate to the beliefs

[19] Though James Otteson, for example, takes it to be paradigmatic that norms emerge via the "impartial spectator procedure," Weinstein's critique of Otteson's "market metaphor" specifically focuses on this as a key point, "The rules of morality are anything but *spontaneous*; they are progressive. Certainly, they are discovered and assented to by different individuals and communities at different times, but this is the product of long-standing inquiry, rational discourse, and the consequences of trial and error. This exploration is dependent on the nature of arts and literature, education and socialization, and a complex account of rationality and historical progress that is far more sophisticated than any economic account of competition of preferences" (2013, 66).

people hold. The model suggests why individuals might experience a disconnect between their social contexts and their moral beliefs, as well as why they might infer from this disconnect that their moral beliefs are transcendently true and therefore not at all dependent on the social conditioning that, in fact, created them. We might say that Smith is ambiguous on the topic of moral absolutes because he never teases these issues out, perhaps in part because, for Smith, these questions were so tied up with religion and natural theology.[20] Less charitably, we might say that he simply fell prey to the very illusion his model predicts.

4.5 Smith's Failure to Account for Change over Time

This computational model strongly confirms the ability of Smith's model to reach moral equilibrium. So strongly, in fact, that it actually raises the opposite problem: after equilibrium is reached, it's not clear how norms ever can change. The sensitivity analyses performed above show how the system will operate differently when variables are initialized at different levels, but in none of the runs does the model exhibit large spontaneous shifts. Given the stability of moral consensus under this model, how are large-scale changes even possible? Otteson has pointed out that "Smith does not consider in any systematic fashion the phenomenon of moral deviancy" (2002, 306). Without a robustly articulated theory of why people might choose alternate norms to adhere to, it is difficult to explain how changes would emerge from the bottom up. Smith has strangely little to say about how norms evolve – strangely, because his underlying theory seems to sketch out principles that should facilitate change, but nothing in *Theory of Moral Sentiments* describes a mechanism that actually brings it about. Axelrod has argued that a "good theory of norms […] should explain three things: how norms arise, how norms are maintained, and how one norm displaces another" (1997, 46). By this rubric, Smith's theory meets the first two requirements but fails the third. This failure is reflected both in the secondary literature and in the performance of the model in simulation. However, Smith is acutely aware of cultural difference, both across geography and across history, and *Theory of Moral Sentiments* seems to be motivated in large

[20] Smith scholars are divided on the extent to which Smith's theory of natural virtues implies a theory of moral absolutes. Macfie (1967, 73) argues that "Smith, like Hume, had no philosophical belief in absolutes, only a 'cool' hope in very gradual, if stumbling improvement." Evensky (2005, 57), however, has argued that, although moral absolutes are not manifest as uniform social norms, they nonetheless subtend social moral jockeying: "Smith's approach to ethics can be understood as moral relativism and invisible absolutes." For Haakonssen (2002, viii), Smith's apparent interest in moral universalism was only "about which features *appeared* to be universal to humanities and which ones *appeared* more or less historically variable." Haakonssen continues, "The universality in question was entirely a matter of observable generality [...] Smith was, in other words, not interested in any metaphysics of morals" (ibid.).

part by his desire to understand why norms differ across place and time. The problem is that he leaves the mechanism of differentiation almost completely implied.

He did leave behind a few tempting clues. In a brief discussion of commercial norms Smith argues that the division of labor results in a division of moral sentiments:

> The objects with which men in the different professions and states of life are conversant, being very different, and habituating them to very different passions, naturally form in them very different characters and manners. We expect in each rank and profession, a degree of those manners, which, experience has taught us, belong to it. (V.ii.3)

Across societies as wholes,

> the style of manners which takes place in any nation, may commonly upon the whole be said to be that which is most suitable to its situation. Hardiness is the character most suitable to the circumstances of a savage; sensibility to those of one who lives in a very civilized society. (V.ii.13)

In this section, Smith seems to veer toward something like cultural relativism. His chapter on custom concludes with a discussion of infanticide, a practice common in ancient Greece, even, to Smith's dismay, "among the polite and civilized Athenians" (V.ii.15). However, such "barbarity" must necessarily be confined to a single practice and can't be representative of the culture as a whole, because, "No society could subsist a moment, in which the usual strain of men's conduct and behaviour was of a piece with the horrible practice" (V.2.16). Taken together, these comments suggest that norms cluster around particular kinds of labor. Professions and nations vary according to the kinds of work expected of them, which sometimes might be the mere effect of chance but other times are driven by the underlying conditions of the labor itself. In any case, norms develop against certain limits that occur naturally: societies might subsist for centuries while countenancing infanticide, but a general permissiveness toward murder would guarantee its immediate demise.

In short, Smith seems to suggest that norms evolve through some process akin to natural selection: norms that are conducive to a community's well-being are more likely to gain assent. If that's true, and if communities change over time as their "circumstances" change, then the evolution of morals should be closely tied to a nation's economic activity. Different circumstances might reward different norms, and the same process that leads the invisible hand to distribute goods and services might also distribute moral sentiments. This possibility brings our attention back to the Adam Smith Problem. Understanding how economic activities affect the development of social norms turns out to be a necessary step to complete Smith's ethics.

5. Modeling the Adam Smith Problem

5.1 Incorporating Economic Behavior in the Model

Moral Markets combines an original approximation of Smith's *Theory of Moral Sentiments* with the trading function of "Sugarscape," the now-canonical virtual economy described in Joshua Epstein's and Robert Axtell's *Growing Artificial Societies* (1996) and adapted to NetLogo by Iain Weaver in 2009.[21] Sugarscape is one of the most studied and re-purposed simulations in the field of agent-based modeling. Its most successful application was in the "Artificial Anasazi" project, which simulated population shifts that occurred among the prehistoric Anasazi Indians in the American Southwest (Dean et al. 2000). Commercial activity in the Sugarscape model involves a simple bartering system. Some agents produce "sugar" and others produce "spice," but all agents need a steady supply of both, and so they must trade to survive. As the supply of either commodity rises and falls, so too its relative price rises and falls. (There is no money in the system, so "price" is the quantity of sugar an agent is willing to trade for a unit of spice, and vice versa.) Describing the model, Epstein and Axtell (1996, 11) specifically invoke Adam Smith's economic theory: "this completely decentralized, distributed achievement of economic equilibrium […] harks back to Adam Smith and the classical economists."

To test how Smith's economic "invisible hand" might affect his moral "golden mean," I borrowed the farming and trading functions of Sugarscape and incorporated them into the Moral Markets model.[22] The functions are largely unchanged; however, I added two key procedures that connect agents' moral and commercial activities. First, I made it possible for agents to specialize by making color a multiplier for production. Spice growers will reap more spice when their outward behavior (B) approaches the value of 140, the purple and pink sections of NetLogo's color scheme. Sugar growers, on the other hand, reap more sugar when their outward behavior approaches 0, where the colors are red and gray. This rule creates an economic incentive toward specialization and, all else equal, implies that agents will tend to deviate morally based on profession. Second, I added a moral component to Sugarscape's trading func-

[21] Iain Weaver, The Sugarscape (2009 NetLogo community model). Weaver's adaptation was itself based on an earlier version by Owen Densmore in 2003. For a detailed description of this adaptation of "Sugarscape," see <http://www2.le.ac.uk/departments/interdisciplinary-science/research/the-sugarscape>, (Accessed January 30, 2018).

[22] The complete Sugarscape model includes many procedures that I did not borrow. For example, Sugarscape simulates a more complex agrarian geography that distributes commodities unequally, while Moral Markets does not. Epstein's model also simulates disruptive forces like epidemics that Moral Markets sets aside in order to focus on questions most directly applicable to the Adam Smith Problem.

tion. Whereas in the original, agents trade on a simply rational calculation of material welfare, in Moral Markets agents are unwilling to do business with a agent whose color differs too greatly from their own. Each agent is assigned a value for their "tolerance," a variable that functions much like "magnanimity." When deciding whether or not to engage in a transaction, each agent calculates the absolute value of the difference between its color and the color of its potential partner. If that difference is greater than their tolerance for difference, they do not perform the transaction.

These two principles, specialization and discrimination, are meant to correspond analogically to Smith's theory of "natural inequalities" in the profits of labor and stock, whereby economic specialization provokes moral demands that many people are unwilling to make. Just as Adam Smith's modest ladies refuse to debase themselves by singing in public, so too agents refuse to do business when their sensibilities are too greatly offended.

5.2 Moral Markets: A Baseline Scenario

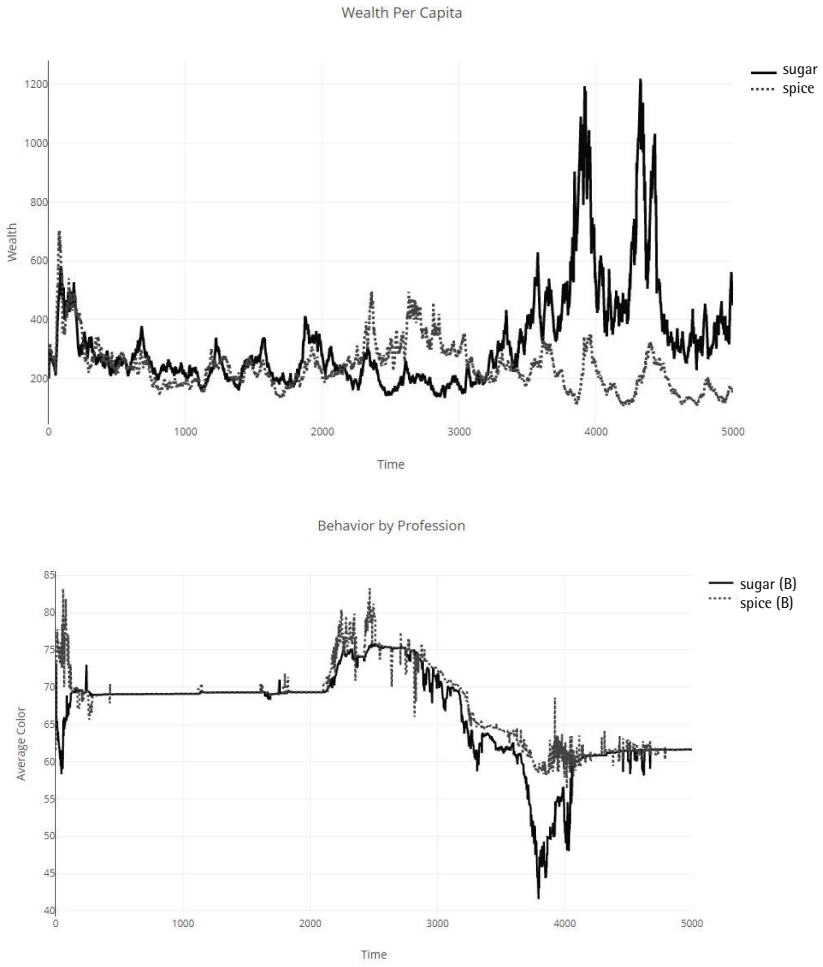
To understand how the model works, it helps to see how agent behavior changes over time and how it varies when economic and moralizing functions are combined. When trading functions are disabled, agent behavior varies according to visibility, tolerance, and self-command, but that variation quickly consolidates around a stable mean. Average wealth and population are, of course, perfectly stable. On the other hand, when trading functions are activated but moralizing functions are disabled, the model achieves a very different kind of stability. Sugar and spice growers quickly specialize: each has an economic incentive to adopt the outward behavior that enjoys the greatest premium on their productivity. As a consequence, the society divides into two species of agents. Because of randomization built into the agents' movements, the relative population and wealth of each type fluctuates, but once they have sorted into behavioral types, such fluctuation occurs around consistently stable averages. When only moral procedures are active, the agents achieve stable norms. When only trading procedures are active, the agents achieve a stable, maximally productive economy.

Taken individually, Smith's moral and economic theories imply societies that tend toward equilibria, but, when his theories are combined, neither equilibrium proves sustainable. Figure 4 illustrates one iteration of the model's performance when both sets of procedures operate in tandem.[23] In this run, agents evaluate each other's colors and adjust their impartial spectators accord-

[23] In this run, the settings were as follows: initial self-command = 80, initial tolerance = 10, vision = 5, and specialization = 10. During the research process, the model was run using many different settings to see how the agents behave under a variety of conditions. Adjusting the parameters can generate many different shapes of distortion.

ingly. However, at the same time, they also gather resources and look for trading partners while refusing to trade with agents whose colors are too different from their own.

Figure 4: Wealth per Capita and Behavior by Profession



Agent behavior by profession over 5,000 steps, showing the relationship between wealth and outward behavior. Between timestamps 2,000 and 3,000, the spice growers benefited from a moral environment that increased their productivity (average behavior > 70). This situation flipped from timestamp 3,000 forward.

The bottom half of Figure 4 shows the average color, or outward behavior, exhibited by sugar and spice growers over time. At initialization, when agents' colors are uniformly distributed between 0 and 140, too many are outside the bounds of their neighbors' tolerance. Sugar and spice growers briefly specialize, but because those new-born specialists are unable to find trading partners, many die. Those that survive quickly conform to a common community standard – the "golden mean" of moral sentiments. (This is typical behavior in the model for the first hundred steps, as the initially randomized agents first begin trading.) For the next 2,000 time-steps, the system reaches near-total moral conformity. Not only is the mean color at the spectrum's midpoint, the standard deviation around that midpoint collapses to zero for long stretches of time with only a few initial blips. Agents reproduce on average every fifty ticks, so 2,000 time-steps represents a period of about forty generations. Analogized to human life-spans, this is roughly the duration of the Dark Ages. Somewhere around time 2,000, this conformity is shattered. The agents' mean color shifts upward to near 80, where it rests for a while before being pulled down dramatically in the steps leading up to time 4,000, then finally settling around 60. This graph illustrates sporadic disunity, as the population shifts wildly from periods of total conformity to bursts of moral diversity.

To understand why these patterns emerged in the agents' moral behavior, it is necessary to see the corresponding changes that occurred in their economic activity. The top graph of Figure 4 shows the comparative wealth agents held over time. For the first 2,000 time-steps, they enjoyed relatively equal prosperity, much like the equality experienced when the economic functions are run in isolation. However, just before time 2,000, the spice growers experienced a brief spike of wealth superiority, freeing them to test behaviors (B) above the mean of 70 that had prevailed to that point. Agents began to conform to this new norm, and spice growers enjoyed a 1,000-tick economic advantage. With communal norms closer to their specialization, the spice growers could be more productive, enjoying greater wealth and enforcing a normative mean closer to their ideal. However, the concentration of wealth among spice growers meant that they were able to support a comparatively larger population of sugar growers, each of whom had an economic incentive to pull the collective moral standards back to their advantage. As the population's wealth and norms reapproached its traditional mean, the whole system radically tipped. A small number of sugar growers were able to capture an enormous quantity of wealth, freeing them to shift in their moral self-display and allowing them to be even more productive as they pulled the entire population in their direction. By time 4,000, the sugar growers' moral and commercial hegemony was secured, and a very small number of agents effectively dominated the entire system. If the model had been allowed to continue, the spice growers might have staged and won another counter-revolution, or the entire population might have died out (as often happens). In any case, this model tells a fable of normative and com-

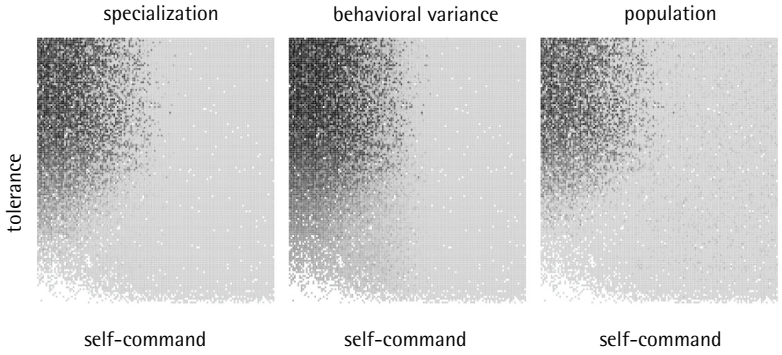
mercial jostling, resulting in long periods of conformity punctuated by moments of economic and moral upheaval.

5.3 Sensitivity Analysis over Key Variables

When a sensitivity analysis is performed across multiple settings these results bear out dramatically. Economic production is closely correlated to the moralizing behaviors of the agents – see Figure 5. The best performing economies are those that are able to support the largest populations. Higher levels of economic productivity are correlated to greater variation in behavior, whether that variation is calculated as a deviation from the mean or as the difference between sugar growers and spice growers. In tolerant societies, producers feel free to pursue their independent, specialized labor. The division of labor increases the total production of the society as a whole, but only when agents are willing to accept variation in normative display. When they are prone to moral jostling, as in the individual run narrated above, economic production is depressed.

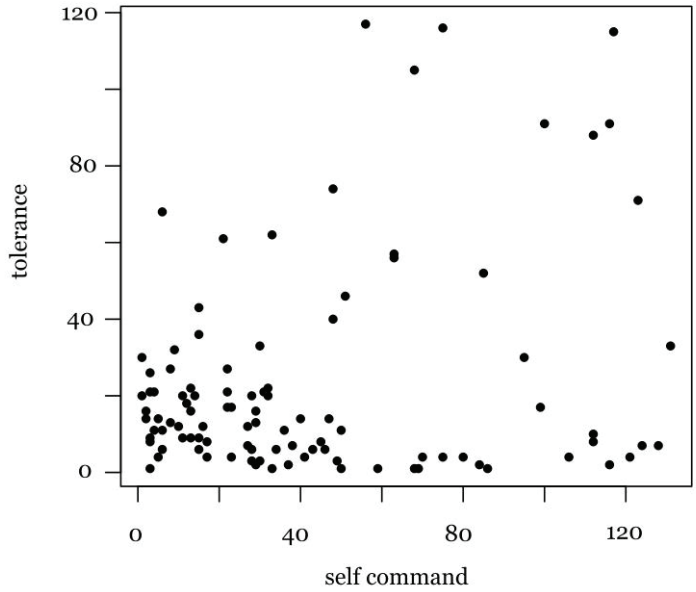
Similarly, inequality correlates negatively to prosperity. In this model, healthy societies tend to distribute their resources more equally. Figure 6 displays the most unequal societies. Runs that demonstrate the most extreme wealth differences between sugar and spice growers tend to appear in the lower-left corner, where agents are willing neither to conform to each other nor to tolerate the differences they encounter. Highly unequal societies tend to exhibit much higher levels of moral conflict, often fluctuating wildly between periods of conformity and periods of fragmentation.

Figure 5: Effects of Tolerance and Self-Command on Agent Performance



Sensitivity testing the correlations between self-command (M, x-axis) and tolerance, showing their combined effect on behavioral specialization (left), behavioral variance (middle), and average population (right). At low levels of self-command and high levels of tolerance, agents are more prosperous (they can support a larger population) while displaying greater variation overall and greater difference between sugar and spice growers.

Figure 6: Tolerance and Self-Command of the Lowest-Performing Economies



Correlation between tolerance and self-command in the 100 runs that displayed the highest level of wealth inequality between spice and sugar growers. Inequality is strongly correlated with a combination of lower self-control and lower tolerance, which tend to depress total economic activity and concentrate wealth among fewer agents.

6. Discussion and Conclusion

The above narration may seem far removed from a close reading of Smith's texts, but it suggests a number of important implications about the Adam Smith Problem.[24] As I see them, these implications can be stated as four general propositions about the ways that "morality mixes with markets," to return to James Otteson's phrase. As interpretations of Smith's philosophy, the first two propositions should not be controversial. The second two might be.

[24] These results were anticipated in outline by Macfie (1967, 111-2) as a likely scenario: "For by their very anonymity, free markets, open to equals, supply the conditions in which the 'exact' rules of justice can operate. History has shown this has considerable truth. Freedom however allows wealth to breed, and this re-establishes inequality and privilege throughout societies. There is then bound to be constant 'justling.' It is inherent in human nature that the struggle for power (his 'place'), is continuous and radical."

6.1 Moral Agreement Is Necessary for Doing Business

This agreement can come about either through general conformity to prevailing social norms or through a general tolerance for behavioral diversity. Under conditions where discrimination is common, market competition operates as a powerful force driving conformity. The internalization of social norms via the "impartial spectator" means that moral beliefs, once firmly entrenched, are likely to persist for long periods of time, even if they reduce overall productivity. However, when moral beliefs are persistently at odds with economic incentives, those beliefs – and, thus, economic activity – are prone to explosive fluctuation.

6.2 How Moral Sentiments Are Structured Affects How Business Is Conducted

In a community where labor is divided, a tolerance for moral specialization will enhance that community's overall production.[25] This is seen in the model when higher levels of economic activity are associated with higher tolerance for moral difference and wider variation in behavior. Smith glimpses this idea in his famous discussion of England's corn laws, where he highlights the economic benefits of morally repugnant actions (in this case, exporting corn for profit while local consumers risk starvation). The converse is also true: general conformity to moral beliefs that don't enhance productivity will tend to depress economic behavior. This should come as no surprise, even though Smith never states it explicitly. Bernard Mandeville's *Fable of the Bees* (1714) famously argued that moral scruples are bad for business. Smith knew Mandeville's work well, and the model diagnoses a similar dynamic precisely. Smith's economic system depends on a willingness to accept the social and moral fragmentation associated with the division of labor; however, the impartial spectator is always pushing with unspecified intensity against such specialization, distorting behaviors and prices.

6.3 Moral Discrimination Causes Inequality

Insofar as members of a community generally conform to social norms that incentivize certain kinds of work and not others, those norms encourage and enforce inequality. This result is seen in the model when agents willingly sacrifice productivity to conform to the norms of their more prosperous customers. Smith seems to recognize this point in his discussion of the "natural inequali-

[25] This view, or something close to it, has been advocated recently by McCloskey (2010), who argues that an eighteenth-century transformation in moral values – placing greater dignity in the merchant class and common trades – was the primary factor driving Britain's economic success.

ties" that attend the "Wages and Profit in the Different Employments of Labour and Stock." He compares butchers to other common trades and inn keepers to farmers, concluding that morally repugnant uses of one's time and one's capital will tend to earn a premium on one's compensation: if the work is "odious," the workers will tend to earn more. Bucking moral consensus to pursue profit does provide individuals with market advantage, the model confirms. However, if those individuals are successful, not just compared to their immediate competitors but compared to the community as a whole, the model predicts they will pull moral consensus in their own direction. They will then earn extra compensation at least in part because of moral beliefs that are widely assented to even though many people who work to support them are excluded from their benefits. These inequalities are not the result of the ebb and flow of supply and demand, at least not in the conventional sense. Long-term structural inequalities are enforced socially and internalized psychologically, and so they can remain rigidly in place over many lifetimes, despite other economic incentives that might push against them. When considered separately, nothing in *Wealth of Nations* or *Theory of Moral Sentiments* predicts a caste system. Considered together, I believe they do.

6.4 Because They Are Spontaneous but Persistent, Moral Sentiments and the Inequalities They Engender Seem Natural, but Are Not

I have already argued that Smith's faith in natural virtues reflects a kind of false consciousness. Because moral beliefs exhibit a stability that individuals' behavior and social norms do not, it is easy to see why Smith was comfortable in taking them to be "natural." So too, inequality. On the one hand, this last proposition veers closely toward a traditional Marxist critique of morality as "ideology." I do mean to endorse such a critique, but I also mean something more narrow and precise. Smith's belief in a "natural price" for commodities and labor is compromised by the artificiality of the norms that structure the rates of supply and demand. If there is no such thing as a natural virtue, there can be no such thing as a natural price, because the price of any good or service is inevitably bound up in the moral jockeying of competing embarrassments. The notion that any good or service has a natural price to which it will tend in the absence of distortion rings false, because there is no theoretically possible absence of distortion. A "natural" price is a price distorted by norms we assent to. Only when they are distorted by norms with which we disagree (nepotism, insider trading, racial discrimination, etc.) do prices seem unnatural. At bottom, price is a function of shame.

In the above discussion, I've hoped to demonstrate that easy-to-overlook details from Smith's philosophy entail significant consequences that Smith himself did not foresee. I have marshaled two distinct kinds of evidence to support

this claim: my reading of Smith's works and a simulation meant to correspond to that reading. Both are liable to objection. Better models of Smith's theory might be developed, and certainly the model could be developed in greater detail. In Moral Markets, homogeneous agents trade opinions about only one "moral sentiment," whereas in Smith's *Theory* people differentiate among each other and among virtues. For example, it would be interesting to see how introducing gender and sexuality might distort the agents' behaviors and beliefs. On the economic side, three notable absences include money, geography, and policy. As written, the model has no function for money, so wealth accumulation is limited and agents can never enter into employment relations. The simulation's treatment of space is similarly abstract. The placement of farms, towns, and trade routes might very well affect the circulation of norms, as well as goods, in measurable and replicable ways. Lastly, the model might include functions that formalize long-dominant norms into laws, allowing to test the relative effects of norms and policies. However, my sense is that any of these improvements would only further strengthen the core argument of this paper: Adam Smith was wrong to treat his subjects separately, and we ought not to persist in his error.

References

Axelrod, Robert. 1997. Promoting norms: An evolutionary approach to norms. In *The complexity of cooperation: Agent-based models of competition and collaboration*, 44-68. Princeton: Princeton University Press.

Becker, Gary. 1957. *The economics of discrimination*. Chicago: University of Chicago Press.

Bonnet, John. 2007. Charting a new aesthetics for history: 3D, scenarios, and the future of the historian's craft. *L'histoire Sociale / Social History* 40 (79): 169-208.

Conte, Rosaria, Giulia Andrighetto, and Marco Campennì, eds. 2014. *Minding norms: Mechanisms and dynamics of social order in agent societies*. Oxford: Oxford University Press.

Dean, Jeffrey S., George J. Gumerman, Joshua M. Epstein , Robert L. Axtell, Alan C. Swedlund, Miles T. Parker, and Steven McCarroll. 2000. Understanding Anasazi culture through agent-based modeling. In *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes*, ed. Timothy A. Kohler and George J. Gumerman, 179-206. Oxford: Oxford University Press.

Dixon, Dan. 2012. Analysis tool or research methodology? Is there an epistemology for patterns? In *Understanding digital humanities*, ed. David Berry, 191-209. New York: Palgrave Macmillan.

Dretske, Fred. 1994. If you can't make one, you don't know how it works. *Midwest Studies In Philosophy* 19 (1): 468-82.

Epstein, Joshua M. 2006. *Generative social science: Studies in agent-based computational modeling*. Princeton: Princeton University Press.
Epstein, Joshua M., and Robert Axtell. 1996. *Growing artificial societies: Social science from the bottom up*. Cambridge, MA: MIT Press.
Evensky, Jerry. 2005. *Adam Smith's moral philosophy: A historical and contemporary perspective on markets, law, ethics, and culture*. Cambridge: Cambridge University Press.
Gavin, Michael. 2014. Agent-based modeling and historical simulation. *Digital Humanities Quarterly* 8 (4).
Gilbert, Nigel. 2008. *Agent-based models*. Los Angeles: SAGE Publications.
Gilbert, Nigel, and Jim Doran. 1994. *Simulating societies: The computer simulation of social phenomena*. London: UCL Press.
Göçmen, Doğan. 2007. *The Adam Smith problem: Human nature and society in The theory of moral sentiments and The wealth of nations*. London: Tauris.
Griswold, Charles. 1999. *Adam Smith and the virtues of enlightenment*. Cambridge: Cambridge University Press.
Haakonssen, Knud, ed. 2002. *Theory of moral sentiments*. Cambridge: Cambridge University Press.
Lloyd, Genevieve. 2013. *Enlightenment shadows*. Oxford: Oxford University Press.
Macfie, Alec L. 1967. *The individual in society: Papers on Adam Smith*. London: Allen & Unwin.
McCloskey, Deirdre N. 2010. *Bourgeois dignity: Why economics can't explain the modern world*. Chicago: University of Chicago Press.
Miller, John H., and Scott E. Page. 2007. *Complex adaptive systems: An introduction to computational models of social life*. Princeton: Princeton University Press.
Otteson, James. 2002. *Adam Smith's marketplace of life*. Cambridge: Cambridge University Press.
Railsback, Steven F., and Volker Grimm. 2008. *Agent-based and individual-based modeling: A practical introduction*. Princeton: Princeton University Press.
Sen, Amartya. 2009. *The idea of justice*. Cambridge, MA: Harvard University Press.
Smith, Adam. 1759 [1982]. *The theory of moral sentiments*, ed. David D. Raphael and Alec L. Macfie. Indianapolis: Liberty Fund.
Smith, Adam. 1776 [1982]. *An inquiry into the nature and causes of the wealth of nations*, ed. Roy H. Campbell and Andrew S. Skinner. Indianapolis: Liberty Fund.
Weinstein, Jack Russell. 2013. *Adam Smith's pluralism: Rationality, education, and the moral sentiments*. New Haven: Yale University Press.
Will, Oliver, and Rainer Hegselsmann. 2014. Modeling Hume's theory of norm emergence. In *Minding norms: Mechanisms and dynamics of social order in agent societies*, ed. Rosaria Conte, Giulia Andrighetto and Marco Campennì, 68-80. Oxford: Oxford University Press.
Winsberg, Eric B. 2010. *Science in the age of computer simulation*. Chicago: University of Chicago Press.