TRUST, TRADE, AND MORAL PROGRESS: HOW MARKET EXCHANGE PROMOTES TRUSTWORTHINESS

By Jonathan Anomaly

Abstract: *Trust is important for a variety of social relationships. Trust facilitates trade, which increases prosperity and induces us to interact with people of different backgrounds on terms that benefit all parties. Trade promotes trustworthiness, which enables us to form meaningful as well as mutually beneficial relationships. In what follows, I argue that when we erect institutions that enhance trust and reward people who are worthy of trust, we create the conditions for a certain kind of moral progress.*

KEY WORDS: trust, trustworthiness, trade, markets, moral progress

I. Introduction

Many people have defended market exchange on the grounds that it increases material prosperity and respects the liberty of buyers and sellers. This defense often rings hollow to communitarians, who emphasize moral goals that they suspect are better achieved by living together in communities where risk is pooled, income is more evenly divided, and social values can be taught and transmitted to members of the group. Of course, communitarian critics of markets are free to forge their own communities in market society, but they are right that none of us can insulate ourselves from the cumulative cultural effects that markets produce. As Samuel Bowles observes, "Because states, communities, and markets influence the process of cultural evolution, any normative evaluation of the role and scope of these institutions must attempt to take their cultural effects into account."[1] Defenders of markets do themselves no favors, then, by ignoring the social orders that markets help shape, and the legal institutions that shape market exchange.

If we could show that the free exchange of goods tends to promote the development of desirable character traits and social norms in addition

[1] Samuel Bowles, "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions," *Journal of Economic Literature* 36 (1998): 105. Deirdre McCloskey concurs: "I do not want to rest the case for capitalism, as some of my fellow economists feel professionally obligated to do, on material achievement alone" (*Bourgeois Virtues*: *Ethics for an Age of Commerce* [Chicago: University of Chicago Press], 22). McCloskey's discussion of the reciprocal relationship between certain character virtues and capitalism is important. She focuses our attention on the astounding leap in Western living standards beginning in the mid-1800s, and proposes that changes in some of the virtues embedded in Western culture were both a cause and consequence of this success.

doi:10.1017/S026505251700022X

to material prosperity, this would be a crucial argument for treating free trade as a moral default. In effect, it would shift the burden of proof onto those who wish to restrict trade.[2] I will argue that in typical cases market interactions tend to promote trustworthiness. I do not assume that it is always better to be trustworthy than untrustworthy. But I do assume that trustworthy people are more likely to be better trading partners, and more likely to enjoy important relationships like love and friendship. Before making this argument, we need an account of trustworthiness, and the relationship between trust and trade.

II. Being Trusted and Being Trustworthy

The concept of trust is familiar enough that we can treat it as a primitive term. In standard cases, trust involves an expectation that a person will perform a particular action.[3] Trust*worthiness* is more complicated. Some use "trustworthy" to describe a character virtue. For example, Deirdre McCloskey describes a trustworthy person as someone who is generous rather than greedy.[4] While this is a legitimate use of the word, I want to explore a sense of trustworthiness that designates a psychological disposition rather than a character virtue (although trustworthiness in the sense I discuss often underlies character virtues).

On David Gauthier's influential account, trustworthiness is "the capacity which enables its possessor to adhere, and to judge that he ought to adhere, to a commitment which he has made, without regard to considerations of advantage."[5] Rather than a *capacity*, I will construe trustworthiness as a *disposition* to adhere, and to judge that we ought to adhere, to our commitments without considerations of advantage.[6] On this view, trustworthiness is one kind of solution to a class of commitment problems in which there are immediate advantages to acting one way, but long-term advantages to being psychologically *disposed* to act in another way.

When there is a trade-off between opportunities to capture small short-run benefits and large long-run benefits, it can pay to constrain our reasoning in ways that make the short-run benefits seem less attractive.

[2] This does not mean there are no good arguments for restricting trade. Problems associated with externalities and public goods often generate at least a *prima facie* case for restricting or regulating trade. See my "Public Goods and Government Action," *Politics, Philosophy, and Economics* 14, no. 2 (2015): 109–128.

[3] Of course, trust is a vague term. Even if we stick with the above definition there are many different *reasons* we trust people, and different *degrees of confidence* that they will perform a particular action. Instead of focusing on trust, I want to rely on a common sense of the word and explore a particular sense of *trustworthiness*, and its relationship to market exchange.

[4] McCloskey, *Bourgeois Virtues*, 159.

[5] David Gauthier, "Morality and Advantage," *The Philosophical Review* 76 (1967): 471.

[6] Emphasizing *dispositions* rather than *capacities* is consistent with Gauthier's focus in *Morals By Agreement* (New York: Oxford University Press, 1986). Trustworthiness, as used in this essay, is a close cousin to what he calls "constrained maximization" in *Morals by Agreement*.

Coercive solutions like "Ulysses contracts" occur when we impose external sanctions on ourselves to help us overcome temptation.[7] Psychological solutions, which include altering our cognitive or behavioral dispositions, are a different way of achieving the same goal. Psychological solutions can be thought of as an implicit contract with our future selves to act in ways that maximize benefits across time. An important difference between coercive and psychological solutions to commitment problems is that in the former case one person empowers *other people* (or external forces) to get him to act in ways that best achieve his long-term goals; but in the latter case a person binds *himself* with psychological tricks that foreclose options that each future self may otherwise find attractive.

Consider the case of a woman with a commitment to remain faithful to her husband. She believes her life will go better if she is faithful, but she is occasionally tempted by the prospect of a fling with a co-worker. She has several strategies available when temptation presents itself. Following Odysseus, she might implore colleagues to blindfold her at corporate cocktail parties, or (more realistically) to interrupt flirtatious conversations with loud banter about her husband and children.[8] Following Prometheus, she might habituate herself to visualize the devastation and betrayal that her spouse would feel if he found out she had an affair.[9] With enough training, her temptation to cheat might weaken. Emotions like love probably evolved to help solve this kind of commitment problem: people who love their partner are less likely to be tempted by the short-run payoffs that seem salient to people who see their partner in purely instrumental terms.[10] One difference between emotions like love, and psychological dispositions like trustworthiness, is that dispositions are more directly under our control.

[7] See Thomas Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960), and Jon Elster, *Ulysses and the Sirens* (Cambridge, UK: Cambridge University Press, 1998). The terminology can be confusing since Ulysses is the Roman name for Odysseus, the mythical Greek character who implored his crew to tie him to the mast and stuff his ears with wax when they neared the island of the Sirens, whose sweet songs lured sailors into dangerous waters.

[8] Legal mechanisms for inducing psychological solutions to the problem of infidelity can include opting for a "fault" rather than "no-fault" divorce in states that offer the choice. This allows potential cheaters to agree in advance to incur a penalty if their philandering breaks up the marriage. Thanks to Allen Buchanan for the example.

[9] I use the Greek god Prometheus (translated, literally, as "foresight") to indicate that psychological solutions can help us overcome myopia.

[10] See Robert Frank, *Passions Within Reason: The Strategic Role of the Emotions* (New York, NY: W. W. Norton & Co, 1998), 54: "A purely rational person who married solely because of exchange possibilities might willingly pledge fidelity, fully aware of all he stands to lose if he reneges . . . The person whose marriage is based on love has an inherent advantage in solving this problem. Love for one's partner imposes an additional cost on the affair, one that is experienced right away. Because the emotional cost of betraying a loved person occurs in the present moment, there is at least some chance it can outweigh the immediate attractions of the affair. The purely rational materialist, who does not experience this immediate cost, will have greater difficulty implementing his pledge."

Cultivating a disposition to keep our commitments can help us overcome weakness of will, affirm our identity and, in some cases, transform us into the person we want to become.[11] A disposition to keep our commitments can even bind together otherwise fleeting future selves in a way that helps us maintain personal identity over time.[12] People who keep their commitments only when they fear they will be caught if they break them are not trustworthy. Instead, trustworthy people *reason* in a distinctive way — they are disposed to act on their commitments without considering the immediate advantages of doing so. Trustworthy people (as I've defined them) believe that they should adhere to commitments because they have made them, not because, when it is time to fulfill commitments, there is value to keeping them.

Consider again the case of a woman faced with an opportunity to cheat on her spouse. According to Robert Frank, "even if the world were to end at midnight, thus eliminating all possibility of penalty for defection, the genuinely trustworthy person would not be motivated to cheat."[13] This is not to say that being trustworthy entails an unwillingness to reassess the reasons for making a commitment. After all, some commitments that we make to ourselves ("I will remain a socialist for the rest of my life") or to others ("I will retaliate to a mild insult with a nuclear strike") are not worth carrying out.

So what do we make of cases in which honoring commitments makes us worse-off on net? The answer is less clear than it may initially appear. One thought is that we should abandon our commitments *whenever* carrying them out makes us worse-off. But this thought precludes us from making genuine commitments, and forces us to forgo the benefits that commitments make possible. A pledge to stay faithful to her husband except when infidelity is likely to lead to better consequences is not a commitment (or is a conditional commitment without any bite), even if it might, under some circumstances, lead to better consequences. The reason commitments can be worth making, and can transform our lives, is that they create opportunities that are otherwise unavailable. If her husband knew that her commitment to him was weak he might never have agreed to marry her, and if he knew her commitments to herself and to other people were precarious, he might never have fallen in love with her. These facts suggest that even when forming or carrying out a commitment in particular cases makes us worse off, being a trustworthy person who is disposed to keep our commitments can make our lives as a whole go better than they would if we lacked this disposition.

[11] Robert Nozick, *The Nature of Rationality* (Princeton, NJ: Princeton University Press, 1994).
[12] Derek Parfit, *Reasons and Persons* (Oxford, UK: Oxford University Press, 1984); and Loren Lomasky, *Persons, Rights, and the Moral Community* (Oxford, UK: Oxford University Press, 1987), 32.
[13] Robert Frank, *Passions Within Reason*, 69.

Although this is plausible as an empirical generalization, there are two obvious problems: first, if we are exceptionally good at disguising our intentions, we might be able to convince other people that we'll carry out our commitments while in fact abandoning them whenever we can get away with it; second, there may be good reasons to abandon our commitments when complying with them makes us worse-off than we would be had we never made them.

The first problem has been recognized for a long time. Plato's example of the Lydian Shepherd who discovers a ring on a corpse that makes him invisible may be the oldest in philosophy to illustrate the problem of opportunistic breach.[14] The Lydian Shepherd uses the ring to seduce the Queen, kill the King, and take over the Kingdom in which he was previously a subject. Plato's brother Glaucon defines justice as keeping our covenants, and argues that only a coward would decline to use the ring to break his commitments and behave unjustly:

> For all men believe in their hearts that injustice is far more profitable to the individual than justice, and he who argues as I have been supposing, will say that they are right. If you could imagine any one obtaining this power of becoming invisible, and never doing any wrong or touching what was another's, he would be thought by the onlookers to be a most wretched idiot, although they would praise him to one another's faces, and keep up appearances with one another from a fear that they too might suffer injustice.[15]

As Robert Frank puts the problem, "if there are genuine advantages in being . . . trustworthy and being perceived as such, there are even greater advantages in appearing to have, but not actually having, these qualities."[16] Since detecting cheaters is enormously valuable, natural selection has enhanced our ability to detect micro-expressions, or involuntary facial cues, that betray the lies cheaters tell.[17] Lying can also be costly to the liar: most people feel guilt or shame for violating vows they make to themselves and other people. Psychopaths are examples of people who exist precisely because detecting cheaters is costly and imperfect, and because there can be material and reproductive benefits to having little or no capacity to feel guilty for tricking people into believing we are trustworthy. While most psychopaths' lives do not go especially well in the modern world, they exemplify the problem of mimicry:

[14] For a nice account of the problem, see David Gauthier, "Three Against Justice: The Foole, the Sensible Knave, and the Lydian Shepherd," *Midwest Studies in Philosophy* 7, no. 1 (1982): 11–29.

[15] Plato, *The Republic*, Book II. http://classics.mit.edu/Plato/republic.3.ii.html.

[16] Frank, *Passions Within Reason*, 9.

[17] Paul Ekman, ed., *Emotion in the Human Face*, 2nd ed. (Cambridge, MA: Malor Books, 2015).

it is sometimes possible to attain the benefits of being trusted without being trustworthy.

Nevertheless, given the challenges of merely *appearing* to be trustworthy, especially in environments in which interactions are repeated and information about cheaters is easily transmitted, it is difficult to thrive as a parasite who seeks trust without being trustworthy. There are, of course, conditions in modern societies that reward parasites capable of masquerading as people who are disposed to keep their commitments. But as I'll argue in the next section, these conditions tend to be a function of background legal institutions, or information asymmetries that can be reduced with legal institutions, not a necessary feature of market exchange.

The second challenge to disposing ourselves to keep our commitments is that it seems to forbid us from considering evidence that keeping our commitments will make us significantly worse-off than we would be if we abandoned them, or never formed them to begin with. On the account developed above, trustworthy people are disposed to adhere, and judge that they ought to adhere, to commitments *without regard to considerations of advantage*. The problem is that sometimes adhering to our commitments will make our lives go much worse than they would if we violated them, so disposing ourselves to stick to our commitments, and to ignore the costs of keeping them, seems irrational.[18]

Consider first an extreme case: keeping our commitment to carry out a nuclear strike in response to a first strike by an enemy. The commitment is only effective if it is believed, and it may only be believed if we first convince ourselves that we have a decisive reason to carry out a nuclear strike if we are attacked first. A commitment involves a belief that we have a reason to act in certain ways in the future. Beliefs, though, are only partly under our control.[19] If our goal is peace through deterrence rather than global destruction, it is not clear whether a *rational* commitment to retaliate with full force is psychologically possible, or whether it requires drugs, self-deception, or a Doomsday Machine programmed to retaliate whatever the consequences.[20]

Assume for the sake of argument that we can dispose ourselves to intend to carry out a retaliatory nuclear strike even though we know that,

[18] Honoring commitments because we have made them, not because of what they'll get us at a particular time, seems inconsistent with the traditional economic assumption that we should never take past investments (or "sunk costs") into account in thinking about how to act. But as Robert Nozick says, "If someone offered us a pill that henceforward would make us people who *never* honored sunk costs, we might be ill advised to accept it; it would deprive us of a valuable tool for getting past temptations of the (future) moment." *The Nature of Rationality* (Princeton, NJ: Princeton University Press, 1993), 23.

[19] This is the main lesson from Gregory Kavka's famous essay, "The Toxin Puzzle," *Analysis* 43, no. 1 (1983): 33–36.

[20] There is a vast literature on this question. The most salient is in Douglas MacLean, ed., *The Security Gamble: Deterrence Dilemmas in the Nuclear Age* (New York, NY: Rowman and Allanheld, 1985).

when the time comes, it would have catastrophic consequences. Gauthier concedes that in this case it would not be rational to carry out our threat even if it were possible to dispose ourselves to *believe* we have reason to carry it out. Abandoning our commitments is rational, Gauthier thinks, when the expected utility of breaking the commitment (at t2) exceeds the expected utility of both forming *and* keeping it (at t1). This proviso allows us to explain the intuition that it is not rational to carry out a deterrent threat even if, assuming it is possible, it may be beneficial to *commit* ourselves to carrying it out.[21]

Although the proviso is plausible, it is incomplete. Once we accept it, we seem to face a dilemma: either we become resolute choosers who ignore welfare-enhancing opportunities that might come from breaking our commitments, or we constantly search for cues in our environment to determine whether keeping our commitments will make us better-off than we would be if we had never made them. In the first case, we become slaves to commitments that our past selves have imposed on our current self; in the second case, we are not reasoning as resolutely as durable commitments require.

An adequate solution to this dilemma should acknowledge two facts: first, it is *costly* to search for information about whether we should abandon our commitments when the benefits of making them fail to materialize; second, even if it is possible to act on the basis of commitments while remaining open to the possibility that breaking them will make our lives go better, it seems psychologically impossible to think this way *too much*.

A plausible theory of trustworthiness will hold that trustworthy people are disposed to ignore the advantages of breaking their commitments, but still sensitive to cues in the environment that violating particular commitments will bring more advantages than making *and* keeping those commitments was expected to bring. The *stakes* of commitments are also important. In low stakes cases we need not be especially sensitive to information about how much better our lives might go if we reconsidered our commitment. For example, if we promise a friend to show up to a birthday party it would be costly, pointless, and even self-defeating to look for evidence that we might get just a little more utility by breaking the commitment and seeing a movie instead. Too much time and energy spent searching for countervailing evidence is both costly and potentially inconsistent with reasoning in a way that commitments require. And the more we do this, the less trustworthy we become as our capacity to reason on the basis of commitments becomes attenuated.

[21] The reason for intending to carry out the threat comes from the fact that, if it is believed, we are more likely to obtain the benefits we seek by issuing the threat (in this case, peace) without having to pay the costs of carrying it out. David Gauthier now believes that the prospect of a failed threat against someone who we have no interest in harming (but whose actions we have an interest in influencing) renders these threats impossible to *rationally* make. See Gauthier, "Assure and Threaten," *Ethics* 104 (1994): 690–721.

The situation is different in high stakes cases, where there is a more obvious trade-off between suffering an attenuated capacity to make commitments, and remaining aware of facts in the particular case that suggest our commitment did not achieve its desired effect. Forming an intention to retaliate to a nuclear strike is illustrative, but the point can be generalized to more commonplace cases, like relationships involving love or friendship. Some people remain stuck in relationships that undermine their career or their happiness because they stubbornly keep their initial commitment to stay with someone who they turned out to be deeply incompatible with. If practical rationality aims at making our lives go as well as possible, it would be irrational to ignore clear evidence that the person to whom we have committed has ambitions or values that are deeply incompatible with ours. But it would not be rational to break our commitment simply because our circumstances have changed or our partner falls ill.

People are trusted for a variety of reasons. But being trustworthy (in the sense I am focusing on) involves a certain way of reasoning, a disposition to ignore the immediate advantages of breaking the commitments we make to ourselves, or to other people. Being trustworthy does not preclude us from being open to evidence that our commitments are making our lives go worse rather than better, but it does preclude us from evaluating the rationality of keeping commitments separately from the rationality of making commitments.

III. Trust and Trade

The reason trustworthiness is important for our purposes is that it can make us dependable trading partners, and can makes our lives go better by opening up important relationships. As Robert Frank notes, "trustworthiness, provded it is recognizable, creates valuable opportunities that would not otherwise be available."[22] In particular, trust lubricates trade by offering assurances to consumers and producers, and *trustworthy* people increase trust to the extent that they can be relied on to keep up their end of the bargain even when it is understood that they might benefit from duplicitous dealing.

The importance of *trust* has long been recognized by economists.[23] Once people understand the benefits of specialization and exchange, they face an assurance problem that can be solved by punishing defectors with external penalties like threats of violence and social ostracism,

[22] Frank, *Passions Within Reason*, 69.

[23] For an overview of the relationship between trust and trade, see chapter 3 of Matt Ridley, *The Rational Optimist* (New York, NY: Harper Perennial, 2011). See also Paul Seabright, *The Company of Strangers: A Natural History of Economic Life* (Princeton, NJ: Princeton University Press, 2010).

or by internal penalties like guilt and shame.[24] It is plausible to suppose shame evolved as a low cost solution to problems that arise when there are benefits to *groups* from following rules, benefits to *individuals* from breaking those rules, and when rule infractions are costly to monitor. Trust emerges and strengthens when we have reason to believe people are likely to follow these rules.

Another way the benefits of trade can be captured through trust is with repeated, dyadic interactions. Game theorists, biologists, and social scientists have long understood that organisms in a game of conflict may find cooperation attractive as long as they believe (or act as if they believe) the game will be played against another particular organism many times.[25]

If we frame trade between two people who have little concern for each other, but who understand the benefits of both trade *and theft*, we get a classic prisoner's dilemma with the payoff structure represented in Figure 1.[26]

Suppose for simplicity that when both try to "take" rather than "trade," a fight ensues over goods that would otherwise be exchanged. In the process calories are dissipated and time is wasted. Both would be better off if they traded peacefully, but each would be best off if the other handed over his goods but didn't fight back. If the interaction only occurs once, the parties appear to be stuck fighting.

But if the prospect of exchange continues far into the future — if the interaction is likely to be repeated an indefinite number of times — rational agents recognize that there are real advantages to testing the waters and seeing whether the benefits of exchange might be captured.[27] In iterated games of conflict, repetition creates crucial incentives that don't exist in single play games.[28] And these incentives can form the foundation for trust.

[24] For more on how social norms and moral emotions can interact to promote mutually beneficial exchange, see Jonathan Anomaly and Geoffrey Brennan, "Social Norms, the Invisible Hand, and the Law," *University of Queensland Law Journal* 33, no. 2 (2014): 263–83.

[25] In game theory, Robert Aumann is widely credited with discovering the conditions under which cooperation can be an equilibrium strategy in an iterated prisoner's dilemma. See "Acceptable Points in General Cooperative *n*- person Games" in Robert Luce and Albert Tucker, *Contributions to the Theory of Games IV* (Princeton, NJ: Princeton University Press, 1959): 287–324. In biology, Robert Trivers appears to be the first to discuss the conditions for conditional cooperation in animals. See "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology* 46, no. 1 (1971): 35–57. In the social sciences, Robert Axelrod discusses similar conditions for human interactions. See *The Evolution of Cooperation* (New York, NY: Basic Books, rev. ed., 2006).

[26] The payoffs presented in this diagram might be misleading after the players cooperate many times and begin to sympathize with one another, thereby transforming the interaction into something more like an assurance game.

[27] The proviso about repeating a game indefinitely is necessary to avoid the problem of backward induction. Other familiar provisos that should be added are that the players have a sufficiently low discount rate, and that they know these facts about one another. Trade may also be modeled as an Assurance Game rather than a PD.

[28] Some evidence suggests trust emerges subconsciously via behavioral cues that make us more or less sympathetic to the specific people we encounter in repeated games. See Robert Frank, "Cooperation through Emotional Commitment," in *Evolution and the Capacity for Commitment*, ed. Rudolph Nesse (New York: Russell Sage Foundation, 2001), 66.

Odin

Thor		Trade	Take
	Trade	8 / 8	0 / 10
	Take	10 / 0	2 / 2

FIGURE 1. The Trader's Dilemma

As Matt Ridley argues, trust and trade almost certainly co-evolved: "It is not at all clear what comes first: the trust instinct or trade . . . plausibly, human beings began tentatively to trade, capturing the benefits of comparative advantage . . . which in turn encouraged natural selection to favour mutant forms of the human mind that were especially capable of trust and empathy — and even then to do so cautiously and suspiciously."[29]

Repeated interactions between particular organisms occur frequently enough in nature that reciprocal altruism is ubiquitous. Among humans in small groups, the same is certainly true. But in modern market societies, many of the exchanges we make are with people we are unlikely to see again. In fact, the supply chain for anything we buy is so long and complex that we can't possibly know even a small fraction of the people we (indirectly) trade with. This would seem to lower the benefits of building a good reputation, and thus reduce trade and encourage cheating.[30] If this were the end of the story, it would pose a serious challenge to the claim that markets encourage trustworthiness.

But there are several ways markets respond to problems created by impersonal exchange.

First, even when we don't have extended and continuous dealings with the same *people*, we often return to the same *firm*. Firms can bear reputations, and act as if they were concerned to maintain their reputation, because they are organized to maximize profits for owners or shareholders. The relevant players desire more income over time, and in a competitive market with property rights that are clear and enforced, they can only increase their income by offering customers what they want at a price they are willing to pay. This provides customers some assurance that when they hand over the cash they'll get the product they pay for.

Second, when uncertainty about the products being exchanged decreases the trust buyers and sellers have in each other, markets for information

[29] Ridley, *The Rational Optimist*, 98.

[30] Samuel Bowles makes this point in "Endogenous Preferences," 93. Adam Smith also recognized that "Where people seldom deal with one another, we find that they are somewhat disposed to cheat, because they can gain more by a smart trick than they can lose by the injury which it does to their character." *Lectures on Jurisprudence* (Oxford, UK: Oxford University Press, 1976 [1763]), 539.

can help solve the problem. Firms compete to provide market participants with information, which reduces the problem that exists when information is either unreliable or asymmetric. The problem is especially prevalent in non-repeated interactions between people who don't know each other. Brand names and reputation are ways firms can increase trust in people who are unlikely to be repeat customers.

Electronic communication and the proliferation of rating agencies has dramatically lowered the cost of producing and transmitting information, which is crucial for brand names to work well. Obvious examples include internet reviews of restaurants and physicians. But perhaps the best example solves a problem made famous by economist George Akerlof.[31] In 1970, Akerlof described the used car market as a paradigm case of market failure created by a lack of trust: since buyers know less than sellers about the condition of the car, buyers hesitate, and markets lead to unexploited gains from trade. This problem has nearly vanished with the emergence of websites that allow customers to check a car owner's maintenance and repair record, and the advent of firms that specialize in buying and selling used cars. In the first case, trust is facilitated directly by information. In the second case, trust occurs because the firm can bear a long-term reputation, and those who run it want to keep their customers happy so they can keep their jobs.

While markets can promote trust, and trust facilitates market exchange, reputation effects are not always enough to prevent bad behavior, especially when the exchange is impersonal. This is where *trustworthiness* comes in.

In economic terms, trust lowers the transaction costs of trade by reducing uncertainty.[32] Many economists have focused on trust for this reason, but trustworthiness can further decrease transaction costs. When people are trustworthy, transaction costs shrink because people can be counted on to keep their commitments even when both parties understand that there are opportunities to capture material rewards by violating them.[33] More specifically, when *sellers* are trustworthy, buyers have additional reasons to believe that a product is as the seller describes; when *buyers* are trustworthy, sellers have good reasons to be confident that buyers will ultimately pay for the products they agree to purchase.[34]

[31] George Akerlof, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84, no. 3 (1970): 488–500.

[32] As Jerry Evensky says, "the weaker our trust, the greater the transaction costs, and the more constrained the market." See "Adam Smith's Essentials: On Trust, Faith, and Free Markets," *Journal of the History of Economic Thought* 33, no. 2 (2011): 251.

[33] We might think *firms* can be trustworthy in the metaphorical sense that they act *as if* they were a person who ignores the immediate advantages of breaking policies to which they have committed themselves. But since I'm using "trustworthy" to refer to a cognitive disposition — a certain way of reasoning — it can only apply to conscious creatures.

[34] The inferred motivations or dispositions that underlie behavior can provide us with additional reasons to trust, which is to say they generate greater confidence that a particular action will be performed. We can imagine someone — or perhaps a robot — who acted

Consider the way in which a trustworthy customer interacts with a trustworthy owner.

When the owner of a sandwich shop turns his back on a trustworthy customer, he can be reasonably sure that the customer will not pull cash out of the tip jar. He cannot be sure of this for a one-time customer with shifty eyes and a suspicious demeanor. But if the interaction is repeated, or he is reasonably good at reading the customer's disposition, the transaction is smooth, and both parties are better off. They may even become friends, and come to find the idea of cheating or lying to each other odious.

Two conditions under which markets are likely to reward trustworthy people are when (a) accurate information can be obtained quickly and cheaply, and (b) competition is robust.

Information is important because for trustworthiness to create opportunities for mutual gain, we need to be able to distinguish who is trustworthy and who is not. According to Bruni and Sugden, "a disposition to be trustworthy is more valuable, and so more worth acquiring, when information can be transmitted through a dense network of associations."[35] As we have seen, although markets can fail when information is incomplete or asymmetric, markets also provide profit-seekers with opportunities to transmit information to those who want it.[36] Market exchange depends on information, but also rewards its provision.

Competition is also important, since it tends to increase the extent to which reputation matters. Competition allows buyers and sellers to avoid cheaters by walking away from an offer and taking their business to more dependable partners. With reliable information and robust competition, markets will tend to reward trustworthy types in two ways: as an *inducement*, and as a *filter*.

By rewarding trustworthy behavior, iterated encounters in markets give us reasons to alter our dispositions. If sellers who keep their commitments are likely to produce happier customers, and this is widely understood, successful business owners may try to change the way they think, and

as if he were trustworthy, but didn't actually have the psychological disposition to act on his commitments without consideration of advantage. This person would be behaviorally indistinguishable from a trustworthy person. But from the fact that we can imagine such a case it doesn't follow that in ordinary circumstances like ours, in which trustworthiness is difficult to fake, the best strategy is always to act *as if* we were trustworthy without being trustworthy, or to think that trustworthiness doesn't matter as much as behavior does. For an empirical test suggesting that inferred mental states can lead to greater mutual confidence that players will reciprocate in iterated trust games, see Kevin McCabe, Mary Rigdon, and Vernon Smith, "Positive Reciprocity and Intentions in Trust Games," *Journal of Economic Behavior and Organization* 55 (2003): 267–75.

[35] Luigino Bruni and Robert Sugden, "Moral Canals: Trust and Social Capital in the Work of Hume, Smith, and Genovesi," *Economics and Philosophy* 16 (2000): 25.

[36] While we are rightly skeptical of information provided by sellers (book jackets and movie posters never feature negative reviews), we can rely on information provided by third parties who have no financial interest in biasing our decisions in ways that make us worse-off, and consequently losing us as customers.

customers may do the same. As Berggren and Jordahl argue, "The economic process of exchange . . . can induce dispositions of trust to emerge, or stimulate conscious decisions to trust others on the expectation that they will be trustworthy and that those who are not can oftentimes be detected."[37] While personal encounters require us to spend time interacting with people before we trust them, impersonal exchanges can increase trust based on the fact that someone is known, by reputation, to be *trustworthy*.

Nevertheless, impersonal exchange may also *limit* incentives to develop a trustworthy disposition. When a customer buys something from a large corporate seller like Amazon.com, the interaction is anonymous, and so does not seem to be the kind of interaction that leads us to alter our dispositions. Corporate executives and boards of directors may decide to create rules that induce trust and brand loyalty, but they are unlikely to cultivate a disposition to reason in a trustworthy way — in the sense discussed in this essay — in response to customer demands. In cases like this, which are fairly common in developed economies, market interactions may be neutral with respect to promoting trustworthiness, even if they promote trust.

In other cases, *untrustworthy* people may thrive because of significant information asymmetries or because of a principal-agent problem. An obvious example occurs when buyers of bonds or stocks don't have a good grasp of how the firm they're investing in works.

Corporate executives like CEOs are sometimes portrayed as snakes in suits who extract resources from shareholders and produce little value in return. This reputation is often undeserved. But there is some reason to worry that executives at large, publicly traded firms have incentives to drive the short-run value of stocks up in ways that promote their own career but not the long-term profitability or productivity of the company they work for. Investors, boards of directors, and executives have different interests (though successful companies devise rules that try to align their interests), and since none are perfectly informed about what the others are up to, executives can sometimes bilk investors. They may even cheat the public at large by getting implicit government bailout guarantees if the company they work for goes bankrupt (as illustrated by the recent failure and subsequent bailout of banks and car companies in the United States).

There are many proposals for correcting information asymmetries and minimizing principal-agent problems, and all of these are best seen as attempts to correct a market failure. But it is important to note that restricting trade by removing competition or abolishing stock markets would likely make the problem worse rather than better. Unlike reliable information in

[37] Niclas Berggren and Henrik Jordahl, "Free to Trust: Economic Freedom and Social Capital," *Kyklos* 59, no. 2 (2006): 161.

the used car market, transparency and information about the activities of corporate executives is hard to produce. Mandatory information disclosures and auditing are imperfect devices for correcting the problem, and it is likely that some such regulations are justified to improve social welfare by diminishing incentives to cheat. But this is a qualification rather than an objection to the claim that market interactions tend to reward companies that can be trusted because they deliver the products they promise, and reward people in repeated interactions for cultivating a trustworthy disposition.

Of course, our propensity to reason in a trustworthy way — to consistently keep our commitments — is not always under our conscious control. It is likely that trustworthiness emerges and is reinforced through habituation, like Aristotle's account of how we acquire character virtues. The fact that psychological dispositions are difficult to cultivate highlights the importance of the fact that market interactions tend to filter out agents who are untrustworthy, at least when the relevant information is accessible and interactions are repeated. They do so, once again, by giving buyers and sellers exit options. They give us the right to say "no." Even if we have little conscious control over our dispositions, as long as untrustworthy people are punished when customers walk away from them and take their business elsewhere, trustworthy types will flourish as the filter of cultural selection works its invisible magic.

Apart from its instrumental benefits, including its ability to minimize transaction costs, trustworthiness as a psychological disposition can be considered a character virtue when it is paired with the right goals and intentions.[38] The fact that trustworthy mobsters might be more effective than untrustworthy ones shows that it is not always good to be a trustworthy person. Commitment to reprehensible ends makes someone more dangerous than he would be if he was willing to reconsider his commitments. But without trustworthiness, and the kinds of commitment that it entails, social relationships would be tenuous and shallow.

Think of trust as the cement of society and trustworthiness as the catalyst of caring relationships like love and friendship. To the extent that markets reward trustworthy people and punish those who routinely violate their commitments, they facilitate the development of a psychological disposition that underlies our capacity to develop and sustain meaningful relationships like friendship and love. The next step is to show which background institutions within which markets operate are likely to maximize trustworthiness.

[38] Phrases like "honor among thieves" suggests it is sensible to say that even those with malicious or otherwise corrupt motivations exhibit character virtues. Aristotle disagrees. I do not wish to make a claim one way or another since it strikes me as a semantic dispute that we can settle by stipulating what we mean.

IV. Institutions and Trust: Optimizing Trustworthiness

> Should the social order be organized to allow moral deviants to gain at the expense of their fellows? Or instead, should the institutional arrangements be constructed in such a way that the immoral actor can gain little, if at all, by his departure from everyday standards of behavior? These questions are based on the acceptance of the idea of progress as applied to social organization . . . [39]
> ~Buchanan and Tullock, *The Calculus of Consent*

Institutions like the rule of law and the enforcement of contracts increase trust by assuring us that commitments are likely to be kept. But they can also promote trustworthiness by creating an environment in which nice guys finish first, rather than last. Unless laws against theft, fraud, and misrepresentation are enforced, cheaters thrive, and information asymmetries allow sellers and buyers to exploit people they interact with.

But how much enforcement should we have? This is a remarkably difficult question to answer because there is sometimes a trade-off between promoting *trust* (in the sense of dependability) and promoting *trustworthiness* (in the sense that one is dependable because one reasons in a particular way). Consider the case of a couple who wants to renovate their house, and is faced with the choice of hiring a carpenter (often called a "contractor") to provide them with a price estimate and then do the job within a budget close to the estimated price. The couple knows a lot less than the carpenter about how to build houses, and carpenters are notorious for underestimating how much money and time a particular job will require. Should we have laws that require full refunds for everything carpenters promise to deliver, but don't?

Perfectly enforced laws that offer full refunds for undelivered services would likely increase trust, but may actually decrease trustworthiness if it disincentivized customers from searching for honest dealers who keep their commitments for the right reasons — that is, regardless of punishment for breaking them. For any particular case, we would like to see full refunds given to customers who do not get what they were promised. But knowing that compensation will be paid in full, no matter what, may mean that producers will be encouraged to honor their contracts only when not doing so will bring legal penalties.

Because of the value of trustworthiness, Erin O'Hara argues that "[t]he challenge of the law is to provide a system for handling breaches of responsibility so that people are encouraged to trade with strangers but not at the same time discouraged from gathering trust-relevant information about each other. That system, moreover, should be one that encourages

[39] James Buchanan and Gordon Tullock, *The Calculus of Consent: Logical Foundations of Constitutional Democracy* (Ann Arbor, MI: University of Michigan Press, 1962), 287.

trustworthy behavior by enabling trustworthy individuals to reap the rewards of their virtue."[40] O'Hara argues that in thinking about the proper institutions that support trade, especially the law of contract, "trustworthiness rather than trust should be maximized" since "the intrinsic motivation to do the right thing is apparently dampened with the imposition of external sanction."[41]

I have argued that trustworthiness is valuable in part because, as a general disposition, it enables us to form important social bonds, and because it lowers the transaction costs of trade and other positive sum interactions. When people are inclined to give up their commitments *whenever* the expected utility of doing so exceeds the expected costs, all of us pay the costs of monitoring, detecting, and potentially punishing them. Relying on legal institutions to enforce agreements can be more costly than relying on psychological dispositions, or on social norms.

Robert Ellickson and Elinor Ostrom have emphasized that in relatively small and homogenous communities with members who regularly encounter one another, social norms often emerge to solve collective action problems, increase trust, and facilitate trade.[42] It may seem puzzling that social norms can do this more effectively than laws backed by formal sanctions. After all, for social norms to effectively regulate behavior, they have to be monitored and enforced. Monitoring deviants and spreading information about them is a public good: the individual who gathers and spreads information about rule infractions incurs the costs of doing so, while the benefits are spread among all community members. But as Richard McAdams has recognized, "[t]he conversation we call 'gossip' is often experienced as a benefit, not a cost, and it usually consists of information about how others have deviated from ordinary behavior."[43]

Our desire for the esteem of others, and the shame and guilt we feel for violating internalized social norms, are social emotions that make policing norms less *psychologically* costly than it would otherwise be. If we enjoy enforcing norms, are rewarded with esteem for doing so, and feel indignant toward those who deviate from them, we will be more willing to police norms than if we were disposed to follow and police social rules only when it is to our immediate and material advantage to do so.

[40] Erin O'Hara, "Trustworthiness and Contract," in Paul Zak, ed., *Moral Markets: The Critical Role of Values in the Economy* (Princeton, NJ: Princeton University Press, 2008), 179.

[41] O'Hara, "Trustworthiness and Contract," 179. It is likely that authors like Erin O'Hara are not using "trustworthiness" in precisely the same way I am, but the way they use the term suggests they agree that trustworthy people reason in ways that differ from people who are merely trusted for reasons having nothing to do with their psychological dispositions or intentions.

[42] Elinor Ostrom, "Collective Action and the Evolution of Social Norms," *Journal of Economic Perspectives* 14, no. 3 (2000): 137–58; Robert Ellickson, *Order without Law*, (Cambridge, MA: Harvard University Press, 1991).

[43] "The Origin, Regulation, and Development of Norms," *Michigan Law Review* 96, (1997), 362.

Thus, the key to social norms working well is the suite of social emotions and psychological dispositions that induce us to monitor and police them, and feel bad about violating them. The desire for esteem and the shame we feel for violating norms can provide sufficient incentives for many of us, much of the time, to follow social norms without considering the advantages of doing so. Herbert Gintis and Samuel Bowles conclude that "where the emotion of shame is common, punishment of antisocial actions will be particularly effective and as a result seldom used. Thus groups in which shame is common can sustain high levels of group cooperation at limited cost . . ."[44]

When social norms direct us to punish people who don't keep their commitments, they can often do so more effectively and at lower social cost than laws and explicit penalties intended to accomplish the same end. And although monitoring deviations from social norms and punishing deviants through social ostracism is more easily accomplished in small groups, it can also be done in large and heterogeneous populations if social norms are internalized and information is easily spread.

This is not to say that legal sanctions are irrelevant when social norms are present; only that they may crowd out the ability of social norms to promote the psychological disposition to honor our commitments without considerations of advantage.[45] By creating strong penalties for all deviations from contracts, a legal system may foster an environment in which we have high levels of trust in the partners we transact with, but relatively low levels of trustworthiness.

Laws and norms can often be considered rival institutions for solving collective action problems and increasing trust. But both can crowd out trustworthiness by laying too heavy a hand on deviants, which might increase short-run assurance but decrease people's long-run propensity to keep their commitments without considering the immediate advantages of doing so.[46] We have seen that trust can promote trade, and vice versa, but we have also seen that creating trust through legal penalties can be socially costly in several ways: legal systems are expensive to operate; they can crowd out norms that sometimes provide similar solutions at lower cost; and they can decrease our tendency to seek out people who are *trustworthy* in the sense that they tend to ignore the advantages of breaking their commitments.

[44] *A Cooperative Species: Human Reciprocity and Its Evolution*, (Princeton, NJ: Princeton University Press, 2011), 193.

[45] Claudia Williamson argues that formal institutions (like legal sanctions) can crowd out informal institutions and rules (like norms and customs), and that informal institutions are a better predictor of economic performance. Williamson, "Informal Institutions Rule," *Public Choice* 139 (2009): 371–87.

[46] One explanation for this rests on the psychological generalization that if people act simply out of fear of the consequences of breaking their commitments, they may fail to dispose themselves to keep their commitments for reasons that are likely to last if the sanctions are removed.

While social norms are emergent orders that cannot be directly imposed by top-down directives, they can be shaped by legal and economic institutions. Contract law should attempt to facilitate trade by increasing trust, but it should do so in a way that induces people to seek trustworthy trading partners. This can be accomplished by giving them some (but not too much) assurance that cheaters will be punished by the state for failing to deliver on their contractual commitments. Social norms may emerge to fill in part of the gap by offering esteem incentives for trustworthy trading partners and ostracism directed at those who are untrustworthy. But laws and norms cannot directly produce trustworthy people.[47] Trustworthy traders are especially likely to emerge in free societies in which competition is robust, and reliable information is easy to share. And while trustworthy traders may not be trustworthy in every dimension of life, or even be good people, institutions that induce people to keep their commitments create the conditions for a certain kind of moral progress. People who are disposed to keep their commitments tend to be better friends and companions, as well as better trading partners.[48]

V. Conclusion

There are clear limits to the kinds of moral progress we can credit markets for creating. Some is speculative. For example, some scholars have pointed out the correlation between economic liberty and toleration, arguing that interacting with different kinds of people for mutual gain in the market tends to make us empathize with people from different backgrounds.[49] Others have argued that the free exchange of goods on terms that are mutually agreeable creates a more pronounced sense of fairness than is found in societies with less exposure to markets.[50]

These are intriguing claims, but I have focused on the extent to which trust can facilitate trade, and trade can promote trustworthiness, a psychological

[47] It is conceivable that people might internalize a norm that directs them to *reason* in a trustworthy way, but norms and laws cannot *force* people to acquire psychological dispositions.

[48] The relevant empirical literature on trustworthiness is scant. But if we could show that most people are capable of radical compartmentalization such that trustworthiness in one domain has no effect on how we reason, or are capable of reasoning, in other domains, this would challenge my thesis. In particular, if we could show that people are *constrained* maximizers in the market but *straightforward* maximizers in personal relationships, this would weaken my conclusion. An empirical result wouldn't automatically undermine the thesis, but it would suggest a more tenuous connection than the one I defend.

[49] Niclas Berggren and Therese Nilsson, "Does Economic Freedom Foster Tolerance?" *Kyklos* 66, no. 2 (2013): 177–207.

[50] Joseph Henrich *et al*, "In Search of *Homo Economicus*: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Association Proceedings and Papers* 91, no. 2 (2001): 73–78. Apart from fairness, Deirdre McCloskey discusses the relationship between markets and character virtues in *Bourgeois Equality*. Jason Brennan reviews the evidence that markets promote rather than diminish important virtues in "Do Markets Corrupt?" Mark White, ed., *Economics and the Virtues* (Oxford, UK: Oxford University Press, 2016).

disposition that serves as the foundation of important social relationships like love and friendship. If the argument is successful, the view that markets corrupt our character and turn us into parasitic capitalists or selfish citizens looks deeply misguided. Instead, the free exchange of goods and ideas can increase our material welfare *and* create some crucial conditions for us to become better people.

Philosophy, Politics, Economics, and Law, University of Arizona