Journal of the History of Economic Thought, 2025

RULES WITHIN OR WITHOUT? ADAM SMITH ON THE ROLE OF DEONTIC EMOTIONS IN DIACHRONIC CONTROL

BY
LAURIE BRÉBAN
AND
LAURENT JAFFRO

The literature on internal commitment cites Adam Smith as a precursor because of his elaboration of diachronic control, and this has given rise to attempts to model his account. Some of these efforts stress the role he assigns to the "general rules of morality" by which the "bulk of mankind" ensure the constancy of their conduct, and interpret them as self-enforcing resolutions. But how could such internal tactics as adopted by weak agents be effective? How could the knowledge of general rules escape self-deception? We take a closer look at what Smith writes about beliefs and emotional dispositions regarding the important rules of morality.

I. INTRODUCTION

Since the 1980s, the literature on intertemporal choice has classified into two categories the devices that individuals implement to overcome diachronic control problems. The first category relates to "external commitment devices" (Schelling 1984). These allow us to control our short-term impulses for the benefit of our long-term interest by engaging in prior action on the external environment (not to mention third-party control)—the paradigm case remaining that of Ulysses tied to the mast at his request to resist the sirens' song (Elster 1984). Conversely, the second category, referred to as "internal commitment devices"—intrapsychic tactics such as personal rules (Ainslie 1992)—

Laurie Bréban: Université Paris 1 Panthéon-Sorbonne, PHARE, F-75013 Paris, France, Email: laurie.breban@univ-paris1.fr; Laurent Jaffro: Université Paris 1 Panthéon-Sorbonne, ISJPS, F-75005 Paris, France, Email: laurent.jaffro@univ-paris1.fr. laurie.breban@univ-paris1.fr

ISSN 1053-8372 print; ISSN 1469-9656 online/25/000001-21
doi:10.1017/S1053837225100746

exclusively relies on the mobilization of internal resources to override short-term impulses at the time they occur.

To illustrate this last kind of device, some economists refer to Adam Smith's *Theory of Moral Sentiments* (hereinafter *TMS*; see, for instance, Schelling 1984). Smith's sophisticated analysis of diachronic self-control has already been widely noted, and many authors have underlined that Smith's approach to "self-command" excludes external commitment devices. Indeed, a careful reader will be struck by the absence of anything resembling Ulysses's technique in *The Theory of Moral Sentiments*. While early modern philosophers who reflected on how to enforce morality, such as Gottfried Wilhelm Leibniz, mixed external and internal devices (Jaffro 2022), it seems that all the commitments Smith is interested in are internal. This aspect has particularly attracted the attention of decision theorists, so much so that the relevance and limitations of their interpretation, which applies the framework of game theory or hyperbolic discounting to Smith's text—at the risk of appearing anachronistic—deserves to be discussed. This will, in turn, allow the richness of Smith's analysis to be highlighted.

According to Stephen Meardon and Andreas Ortmann (1996), Ignacio Palacios-Huerta (2003), or Roland Bénabou and Jean Tirole (2004), Smith's self-command would typically mobilize internal commitment devices. Interestingly, Bénabou and Tirole consider that these devices rely on the general rules of morality, which would act, in *The Theory of Moral Sentiments*, as self-enforcing resolutions or, in their words, as "personal rules." Yet if Smith believes that the general rules of morality do play a crucial role in solving problems of self-control, it is not because they would constitute promises to oneself or resolutions. The question of the conditions for the effectiveness of internal commitments therefore remains open.

Smith was indeed aware of the limits of such internal devices to mobilize the will—limits already remarked upon by previous philosophers such as Thomas Hobbes and Joseph Butler. According to Smith, the efficacy of moral rules comes precisely from the fact that they are not reducible to *personal* resolutions. They are built on the basis of our social interactions and constitute a kind of summary of our past observations about others' behavior and also about the way others assess our behavior and their own. This social experience is internalized in the form of resolutions to do (respectively, avoid) what is the object of "universal" approval (respectively, disapproval) (see *TMS* III.4.7–11). But this internalization is not sufficient to explain the ability of general rules to direct our behavior, as we will argue. According to Smith, the influence of general rules comes from the belief that they are external to ourselves, and indeed often also from their sacralization, that is, the belief that they have a divine origin. As a result, they are the object of specific emotions, which we term *deontic*—with reference to Smith's understanding of duty—that reflect our reverence for them. And it is these emotions that, according to Smith, sustain the will of most individuals so that they can command their passions.

Yet this is precisely what is not captured by the approaches of Meardon and Ortmann (1996) or Bénabou and Tirole (2004), which focus on the internal resources of the individual, the effectiveness of which Smith questions. Smith, in fact, believes that there is a common condition of weakness that makes the effectiveness of personal rules doubtful on the part of agents who, in reality, do not have a very high degree of

self-command. Their approaches explain that individuals manage to mobilize their willpower by maximizing their long-term gains or utility. However, Smith emphasizes the inability of future events to effectively motivate the will, due to their temporal distance, which reduces their emotional intensity. Deontic emotions, on the other hand, are experienced at the very moment individuals engage in intertemporal decision-making, and it is these emotions, according to Smith, that counterbalance the attraction of immediate pleasures.

In the first part of the paper, we highlight Smith's skepticism about internal self-control devices by presenting the limitations he identifies in his theory of the impartial spectator, and which the general rules are supposed to remedy. Thus, we stress the contrast between the use of the rules and the use of the internalized spectator, contrary to certain readings (for instance, Konow 2012). The aim of the second part is to assess the relevance and limits of Smith's concept of general rules. We credit him with a more plausible account—the mixed view—which supplements rules with habitual implementation, affective incentives, and beliefs about their special authority. What is at stake is not just a particular interpretive issue in Smith's thought but a discussion of the prevalent framework within which the problem of the practical force of rules is to be addressed. To this end, we discuss various models of diachronic self-control.

II. WHY RULES ARE NECESSARY AND SPECTATORS INSUFFICIENT

There are at least two passages of *The Theory of Moral Sentiments* in which Smith alludes to what we would recognize today as the issue of diachronic control. The first one makes the solution to diachronic control problems rest entirely on the consideration of the internal impartial spectator and constitutes a typical case of an internal commitment device. The adoption of the point of view of the impartial spectator is often seen in the literature on Smith as the principle on which moral actions are based. However, in the other passage in which he refers to an intertemporal choice problem, Smith insists on an alternative device, the general rules of morality, which he seems to think more effective in mobilizing the will of most human beings, for they are said to direct the "coarse clay of which the bulk of mankind are formed" (TMS III.5.1).

One might object that the second perspective is not really different from the first. After all, general rules do depend on spectatorial evaluations: they are summaries of them. True enough. But this does not change the fact that the function of general rules is to obviate the need for spectatorial evaluation. The dependence of the rules on spectatorial evaluations is therefore genetic, not epistemic. Once a general rule of morality is in place, there is no need to consult the spectator. And when you are following a general rule, you are not consulting a spectator.

In this section, we pay attention to the move from the first to the second perspective.

The Atemporal Influence of the Ideal Spectator

One of Smith's principal allusions to intertemporal choice in *The Theory of Moral Sentiments* occurs when he discusses one of the main virtues of his moral philosophy: "prudence," a self-regarding virtue aiming at our own happiness. On this occasion,

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

Smith explains that prudence consists in the combination of two qualities: (i) "superior reason and understanding, by which we are capable of discerning the remote consequences of all our actions, and of foreseeing the advantage or detriment which is likely to result from them"; and (ii) "self-command, by which we are enabled to abstain from present pleasure or to endure present pain, in order to obtain a greater pleasure or to avoid a greater pain in some future time" (*TMS* IV.2.6). Prudence thus takes the form of a trade-off between one's short-term interest and one's greater long-term interest in favor of the latter.

Self-command plays a key role in the achievement of this trade-off, and in Smith's moral philosophy more generally. It is described as the quality to which the other important virtues of prudence, justice, and benevolence owe their merit: "Self-command is not only itself a great virtue," Smith writes, "but from it all the other virtues seem to derive their principal lustre" (*TMS* VI.3.11). This "luster" arises from the internal struggle that the practice of virtue requires to overcome: "To act according to the dictates of prudence, of justice, and proper beneficence, seems to have no great merit where there is no temptation to do otherwise" (*TMS* VI.3.11).

Thomas Schelling (1984) thinks that Smith means by "self-command" a virtue. To be precise, Smith uses the term to refer to a quality that is susceptible to degree, not always to the very high, virtuous degree. Schelling's own understanding of "anticipatory self-command" is different: it refers to a set of devices that are useful when habitual self-command is lacking: "[I]n my usage, self-command is what you may not need to employ if you already have enough of what Adam Smith meant by it. You don't need the skillful exercise of self-command to cope with shifting preferences if you've already got your preferences under control" (Schelling 1984, pp. 3–4).

In *The Theory of Moral Sentiments*, Smith refers to self-command very early immediately after offering an explanation of the nature of the moral judgments we make about the behavior of others (see *TMS* I.1.1.). According to him, the approval of a spectator stems from the harmony between the passion expressed by the person being observed and the feeling that the spectator himself would have experienced if he had been in the same situation.

Smith claims that this harmony is based not only on the effort of the spectator to adopt the perspective of the person he is judging but also on the effort of that person to moderate her emotion so that the spectator can sympathize with her. It is this latter effort that is the foundation of the virtue of self-command.[1] However, Smith insists that this effort is not necessarily seen as virtuous in itself. It is only considered virtuous when it is challenging:

> There is … a considerable difference between virtue and mere propriety…. Upon many occasions, to act with the most perfect propriety, requires no more than that common and ordinary degree of sensibility or self-command which the most worthless of mankind are possest of, and sometimes even that degree is not necessary. Thus, to give a very low instance, to eat when we are hungry, is certainly, upon ordinary occasions, perfectly

[1] More generally, Smith refers to the virtues linked to the effort of the person principally concerned to moderate her emotions as the "respectable virtues" as opposed to "amiable virtues," which designates the virtues of the spectator who tries to adopt the circumstances of the former (see *TMS* I.1.5).

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

> right and proper, and cannot miss being approved of as such by everybody. Nothing, however, could be more absurd than to say it was virtuous. (*TMS* I.1.1.7)

Interestingly, it is in the context of this first discussion on self-command that the first mention of the impartial spectator appears (see *TMS* I.1.1.7). As we know, the reference to the impartial spectator is essential to the characterization of virtue.

This spectator is not a real person involved in a bilateral interaction, as previously described, but rather an imaginary spectator who embodies the judgment of others that we have internalized through social interactions. Unlike the real spectator, this one has "no particular connection" with us (*TMS* III.3.3). It is in this sense that he is impartial and thus enables us to judge ourselves.

This is why, as *The Theory of Moral Sentiments* progresses, the virtue of self-command is presented as the disposition that enables the overcoming of a confrontation between two alternative points of view that agents can adopt on their situation: what he sometimes calls their "natural view," and the impartial spectator's view (*TMS* III.3.28).[2] These two viewpoints call upon two contradictory motives, leading to opposing behavior:

- The natural view is urged by our first impulses and tempts us to indulge in our passions. It gives a disproportionate picture of one's own situation.
- The (ideal, that is, internal)[3] impartial spectator's view, which reflects our concern for others' judgments, arises from our regard for "dignity" and "esteem." It is only from this perspective, he writes, "that we can ever see what relates to ourselves in its proper shape and dimensions" (*TMS* III.3.1). This standpoint prompts us to command our passions so as to render them appropriate to the object that elicits them.

Self-command leads to the overcoming of our natural perspective by that of the (ideal) impartial spectator. On many occasions, *The Theory of Moral Sentiments* describes the struggle between the natural and the ideal spectator's viewpoints as a trade-off between opposite sorts of interest. As we have seen, in the case of prudence, this takes the form of a trade-off between one's short-term interest supported by the natural view and one's greater long-term interest supported by the impartial spectator's view.[4] From this perspective, it is tempting to interpret Smith's concept of self-command in the light of the "multiple selves" models of intertemporal choice, as do Meardon and Ortmann (1996) and Palacios-Huerta (2003). What is specific to this construal of Smith's analysis in both cases is that they aim at explaining how individuals achieve to command their long-term interest even though they are subject to short-run impulses that prompt them to move away from it.

Stephen Meardon and Andreas Ortmann's representation takes the form of a simultaneous game of the prisoner's dilemma type. The game, inspired by a passage in *The*

[2] On the confrontation between the two alternative points of view that individuals can adopt in a situation, see Bréban (2014).

[3] *Ideal* not in the sense of being perfect, but of existing in one's imagination, thus as opposed to real impartial spectators, that is, other people not party to the case.

[4] For virtues such as beneficence and justice (others-regarding virtues), it takes the form of a trade-off between one's own interest supported by the natural standpoint and another—moral or physical—person's equal or greater interest supported by the (ideal) impartial spectator's perspective, the corresponding vices being exemplified by the sentiments of a European on hearing the news of the destruction of the "great empire of China" (*TMS* III.3.4).

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

Theory of Moral Sentiments,[5] is said to be "intrapersonal" because it involves the interaction of two players, each corresponding to two successive facets of the same individual. The authors call them the "Man Yesterday," who must choose between acting properly or improperly; and the "Man Today," who evaluates the actions of the Man Yesterday.

The use of a simultaneous game is justified, according to the authors, by the fact that when the Man Yesterday is in the grip of passion, he anticipates the judgment of the Man Today—once the passion has passed. In this approach, the temporal conflict (between short-term and long-term interests) manifests itself through the choice of the Man Yesterday, whose "act improperly" strategy is weakly dominant but which the indefinite repetition of the game leads to his abandoning in favor of the "act properly" strategy—under the assumption that each player uses a trigger strategy.[6] Thus, the indefinite repetition of the game leads to the cooperation of the two selves of the individual who had no interest in cooperating in the stage (or base) game. This cooperation between the two selves of the individual is supposed to represent the acquisition of self-command.

Palacios-Huerta (2003), for his part, adopts an intertemporal utility model with a hyperbolic discounting factor to represent Adam Smith's conception of self-command. In such models, hyperbolic discounting expresses the fact that agents' preference for the present decreases over time so that, for example, they prefer the larger and later of two sums of money when both are far away, but they switch to the smaller and earlier when they come close to it. Such a phenomenon corresponds to time inconsistency. Formally, it comes from the violation of the axiom of stationarity,[7] dear to the exponential discounting model, which is challenged by hyperbolic discounting. According to Palacios-Huerta (2003), the Smithian individual manages to overcome the attraction for what is immediate—short-run impulses—thanks to the present influence of past and anticipated experiences, as shown by the recursivity of the utility function representing preferences.

This interpretation is in line with Meardon and Ortmann's (1996) idea that it is the present influence of distant events that would, according to Smith, enable individuals to control their passions and act in accordance with their long-term interest. Moreover, both

[5] See *TMS* III.4.4.

[6] According to the folk theorem, in an infinitely repeated game, there is a multitude of subgame perfect equilibria—that is, a multitude of combinations of actions that constitute for each player a best response to a combination of actions of the other player—several of which may be Pareto optimal. However, in order to simplify their presentation, Meardon and Ortmann (1996) assume for each player the existence of trigger strategies. Indeed, in an infinitely repeated game, the actions of the current period are chosen conditional on the observations made on the behavior of the other player in the previous period. This gives each player the possibility to punish the other if he deviates from the collective course of action defined at the beginning—according to a criterion external to the game—by adopting actions that only grant him a payment lower than his minimax in the base game. In Meardon and Ortmann's model, the Man Today's trigger strategy is the one we are most interested in, since only the Man Yesterday has an incentive to deviate from the "act properly and evaluate routinely" equilibrium. This strategy consists of routinely evaluating the Man Yesterday's actions —i.e., effortlessly—as long as he chooses to act properly, and punishing him by choosing a serious evaluation of his actions, likely to lead to a feeling of being blameworthy, throughout the rest of the game —i.e., indefinitely—if he chooses to act improperly even once.

[7] According to this axiom, if two alternatives x and y are indifferent at some dates t and $t + \tau$, they are also indifferent at any other dates s and $s + \tau$.

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

representations are based on the assumption that individuals maximize their long-term gains or utility.

Smith insists, however, on the very weak influence of temporally distant events on present decisions and on the difficulty of individuals to adopt a long-term perspective. It is not the consideration of the greater pleasure, he claims—in other words, our long-term interest—that prompts us to command our passions. To the spectator,

> the pleasure which we are to enjoy ten years hence interests us so little in comparison with that which we may enjoy today, the passion which the first excites, is naturally so weak in comparison with that violent emotion which the second is apt to give occasion to, that the one could never be any balance to the other, unless it was supported by the sense of propriety. (*TMS* IV.2.8)

Here, Smith has in mind the process by which the impartial spectator's view overcomes the natural view, in the struggle that opposes them, each standpoint expressing distinct temporal inclinations. Smith assumes that the temporal distance of pleasure and pain reduces the intensity of their perception, so that they would give rise to less and less violent passions. This would result in a natural tendency to overvalue the present compared to the future, so that a great forthcoming advantage would not be able to compensate for a lesser present one; this is very similar to what we usually called "impatience." after Peter Fishburn and Ariel Rubinstein (1982). "Impatience" expresses formally that if something pleases us, we always prefer it sooner rather than later. In Smith's context, only a present gratification, the one associated with "the consciousness that we merited the esteem and approbation of everybody" (*TMS* IV.2.8) when favoring future over the present, is able to balance the consequences of our impatience. Now, such a gratification is the result of our endorsing the impartial spectator's non-impatience. To the spectator, reviewing the behavior of the people concerned,

> their present, and what is likely to be their future situation, are very nearly the same: he sees them nearly at the same distance, and is affected by them very nearly in the same manner. He knows, however, that to the persons principally concerned, they are very far from being the same, and that they naturally affect them in a very different manner. He cannot therefore but approve, and even applaud, that proper exertion of self-command, which enables them to act as if their present and their future situation affected them nearly in the same manner in which they affect him. (*TMS* VI.1.11)

This shows that taking into account the impartial spectator point of view does not mean that we now share his preferences over time: it is then the "sense of propriety," not the consideration of the greater advantage to come, that leads individuals to command their passions and to follow their long-term interest. By adopting the impartial spectator's perspective, they are conducted to act *as if* they were indifferent between present and future, *as if* their temporal preferences had changed—or, more formally, *as if* they were no more subject to "impatience."[8] However, Smith emphasizes that this is nothing more than an *as if.* Although self-control appears to take the form of an intertemporal trade-off, this is only on the surface. People actually make a trade-off between two kinds of present

[8] Such a behavior may be described by the "condition of patience" that Bianchini (2016) uses in order to characterize "temperate" preferences, à la James Mill.

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

pleasures: the one arising from the present object of choice and the one arising from the sentiment of being praiseworthy. Thus, intertemporal choice here comes down to a synchronic trade-off.

Now, Smith expresses skepticism regarding the influence of the impartial spectator and the strength of the sense of propriety: "None but those of the happiest mould are capable of … acting upon all occasions with the most delicate and accurate propriety" (*TMS* III.5.1). Most of the time, representations one has of the impartial spectator's view tend to be altered and this has temporal effects.

Effective Diachronic Control and the Need for General Rules of Morality

Smith's skepticism regarding the influence of the impartial spectator is expressed in the second passage where he refers to the diachronic control issue. To show how the device of the impartial spectator may fail, he analyzes the way we are likely to examine our situation at two successive dates: when we are about to act and after we have acted. Such an analysis comes along with illustrations in which two alternative options, respectively dictated by the natural and the impartial spectator's views, are evaluated: (i) to succumb to passion or (ii) not to succumb. Interestingly, in these illustrations, the evaluation of neither option entails—at least explicitly— an intertemporal trade-off. Most of them depict an agent having to examine such actions as (i) injuring someone by revenge or, alternatively, (ii) abstaining from revenge. Let us review how Smith accounts for our evaluation at each date.

Smith claims that initially, when we are about to act, passion tends to "discolor" our representation of the impartial spectator's point of view (*TMS* III.4.3). In the struggle that sets the impartial spectator in opposition to the natural point of view, this results in the victory of the latter, which thus leads us to depart from what the sense of propriety commends to us. Smith characterizes as one form of "self-deceit" the process by which we tend to ascribe to the impartial spectator the natural view (*TMS* III.4.3). A different form of self-deception consists in avoiding the gaze of the spectator or refusing to endorse it (*TMS* III.4.5–6). When the action is over, Smith claims, we are in a better disposition to enter the sentiments of the impartial spectator and to revise our evaluation of our circumstances:

> The man of to-day is no longer agitated by the same passions which distracted the man of yesterday: and when the paroxysm of emotion, in the same manner as when the paroxysm of distress, is fairly over, we can identify ourselves, as it were, with the ideal man within the breast, and, in our own character, view, as in the one case, our own situation, so in the other, our own conduct, with the severe eyes of the most impartial spectator. (TMS III.4.4)

Such a change of perspective may give rise to time inconsistencies. Faced with the choice between two simultaneous options x (succumb to passion) and y (not to succumb), an agent may prefer x to y at date t_0, y to x (regret) a bit later in t_1, and, still later, at t_2 again x to y.

However, note that even in circumstances that are considered as favorable to the adoption of the impartial spectator's view, Smith remains quite skeptical regarding its influence, so that such time inconsistencies may be infrequent: "It is seldom," he claims, that our judgments about our character are "quite candid," even when the action is over

(*TMS* III.4.4). This leads us to "purposely turn away" from the impartial spectator's perspective, by endeavoring to "exasperate anew those unjust passions which had formerly misled us" (*TMS* III.4.4). Thus, we restore a form of subjective consistency by intentionally rejecting the viewpoint of the spectator.

People also tend to persist in the form of self-deceit that consists in an alteration of the impartial spectator's view, and through this twist they remain time-consistent. So contrary to what is asserted by Meardon and Ortmann (1996) and Palacios-Huerta (2003), going against what the impartial spectator dictates does not necessarily result in temporal inconsistency. Smith considers that in some cases an individual may consistently engage in an improper action through self-deceit.

Now, one of Smith's proposals that has gone largely unnoticed among economists—with the exception of Meardon and Ortmann (1996) and Bénabou and Tirole (2004)—probably because the spectator theory has attracted so much attention, concerns the role of general rules.[9] Reliance on the spectator cannot suffice to counteract a tendency to self-deception (Konow 2012). This is one of the main reasons why Smith is interested in the regulatory power of the general rules of morality that are the content of what he calls the "sense of duty." He considers that for most people, the sense of duty—which prompts them to respect moral rules—is more effective, and thus more important, than a sense of propriety they most often lack: "Many men behave very decently, and through the whole of their lives avoid any considerable degree of blame, who yet, perhaps, never felt the sentiment upon the propriety of which we found our approbation of their conduct, but acted merely from a regard to what they saw were the established rules of behaviour" (*TMS* III.5.1).

But we should not conclude from these lines that Smith is not a sentimentalist. Even if the sense of duty is clearly distinct from the sense of propriety, even if most people are more likely to follow general rules than the ideal spectator, the fact remains that Smith's theory of value judgment is based on a sentimentalist response theory, according to which x has the value that fair and well-informed spectators assign to it through their affective responses (Jaffro 2024).

Neither do these lines lead to questioning the centrality of the sympathetic process for Smith. Sympathetic interactions are fundamental to explaining the emergence of general rules. Moreover, general rules make it possible to reach the degree of self-command necessary for the communication of affects to take place—the self-command necessary to reach, for instance, the point of propriety. However, they allow for the qualification of the influence generally attributed to the impartial spectator in *The Theory of Moral Sentiments*. This nuance is provided by Smith himself in the chapter immediately following the one devoted to the impartial spectator: "Of the Nature of Self-Deceit, and of the Origin and Use of General Rules" (*TMS* III, 4). For Smith, the impartial spectator cannot explain how most people manage to act properly because of self-deception. Only the wise manage to mobilize it effectively. In a nutshell, more or less as David Hume is a global sentimentalist with a locally non-sentimentalist account of conventions in book 3 of his *Treatise*, so Smith is a global sentimentalist with a locally non-sentimentalist or not obviously sentimentalist account of "duty" and rules.

[9] There is relatively more work on Smith's general rules among political scientists and philosophers (Remow 2007; Fricke 2011; Schliesser 2017; Liu and Weingast 2020; Montes 2020).

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

III. WHY RULES ARE INSUFFICIENT PER SE AND HOW TO REINFORCE THEM

It remains to be seen how these rules can acquire effective authority over practice and why they lack it without some reinforcement. According to Smith, in contrast with the device represented by the impartial spectator, general rules introduce a temporal dimension to the decision to control one's impulses by allowing foresight of the future consequences of one's actions. This may suggest that respect for them is based on the fear of regret. However, as shown in this section, this kind of explanation is not sufficient to account for the efficiency of rules. Let us first pay attention to the epistemological status of general rules of morality.

The Epistemology of General Rules: From Summaries to Norms

General rules offer us reliable information regarding the "sentiments of mankind" with respect to our passions, information that is possible to appeal to at the time they occur. The case of the man of furious resentment illustrates this:

> Unless his education has been very singular, he has laid it down to himself as an inviolable rule, to abstain from [revenges] upon all occasions. This rule preserves its authority with him, and renders him incapable of being guilty of such a violence. Yet the fury of his own temper may be such, that had this been the first time in which he considered such an action, he would undoubtedly have determined it to be quite just and proper, and what every impartial spectator would approve of. But that reverence for the rule which past experience has impressed upon him, checks the impetuosity of his passion, and helps him to correct the too partial views which self-love might otherwise suggest, of what was proper to be done in his situation. (*TMS* III. 4.12)

Although general rules constitute norms of action, they are based directly on the observation of empirical regularities. At root, Smithian rules exemplify what John Rawls called the "summary view" of rules (Rawls 1955): "The general maxims of morality are formed, like all other general maxims, from experience and induction. We observe, in a great variety of particular cases, what pleases or displeases our moral faculties, what these approve or disapprove of; and by induction from this experience we establish those general rules" (*TMS* VII.3.2.6).

To highlight the conditional nature of these general rules or maxims—in this discussion Smith makes no distinction between the two—that is, the dependence of the observed regularity on factual or counterfactual circumstances, Smith gives as an example our attitude toward the grief of a stranger who, we learn, has just lost a loved one. For a variety of contingent reasons, Smith explains, we are unable, under the circumstances, to actually have the feelings that might correspond to those of the person concerned. This prevents us from judging the propriety of that expression of grief. We run the risk of showing a completely inappropriate indifference. Fortunately, we "have learned from experience" what attitudes are in such circumstances considered appropriate and are expected respectively from the person concerned and the bystander:

> It is upon the consciousness of this conditional sympathy, that our approbation of his sorrow is founded, even in those cases in which that sympathy does not actually take

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

> place; and the general rules derived from our preceding experience of what our sentiments would commonly correspond with, correct, upon this, as upon many other occasions, the impropriety of our present emotions. (*TMS* I.1.3.4)

Sympathy here is said to be conditional because it is counterfactual. It is not the case that there is a correspondence of feelings, but we know that this would be the case if we were prepared to have feelings in relation to the information we have about the situation of the person concerned. This knowledge, based on common experience, is general, and allows the formulation of a conditional rule: if placed in circumstances *X*, people tend to have such-and-such attitudes and behave in such-and-such ways. In Smith's example involving third-party evaluation, the attitudes concerned are also those of the spectator, and relate to politeness.

Smith describes in a very concise way the transformation of descriptive rules from repeated observations into personal rules of conduct: "Our continual observations upon the conduct of others insensibly lead us to form to ourselves certain general rules concerning what is fit and proper either to be done or to be avoided" (*TMS* III.4.7). However, there is no fallacious transition from "is" to "ought," provided that the observance of the maxim is instrumental to a desire previously given. For a general observation that people tend to have such-and-such an attitude or to behave this way rather than that way in such-and-such circumstances gives rise to a normative maxim only if the empirical observation is accompanied by a desire to "regulate" one's conduct in those circumstances, that is, in the case of the rules of morality, by the desire for virtue and the aversion to vice.

Since the material for the process of induction by which all general rules are formed is provided by repeated instances of particular evaluative operations, it is impossible to account for these evaluations as just being the application of general rules to particular cases, as a certain moral rationalism would have it (*TMS* III.4.8). The evaluation is based on the exercise of the sense of propriety, that is, basically, on a correspondence of feelings. Of course, this does not prevent the rules from constituting evaluation criteria dependent on the actual use of the spectator, that is, "standards of judgment" (*TMS* III.4.11), precisely because they are recapitulations of already established evaluations.

These standards would have little authority if they only reflected isolated personal experience. The moral experience that is at the root of the formation of general rules is not only personal but also social, as shown again by the case of the man of furious resentment whose education and "observations upon the conduct of others have taught him how horrible all such sanguinary revenges appear" (*TMS* III.4.12). It is important to note that the similarity of everyone's experience leads to maxims that are "universally acknowledged and established, by the concurring sentiments of mankind" (*TMS* III.4.12).

But the central practical function of general rules is motivational, since they allow for a "correction" of feelings. This is where the main difficulty lies. How can empirical generalizations, which describe what our experience is, influence action or passion? There is a mystery here that will be cleared up in the next section, when we look at the roles played by desires, feelings, and habits in the implementation of these rules.

Smith admits that, since maxims are formed by induction, and since induction is considered an operation of reason (it is indeed a reasoning about matters of fact), "reason is undoubtedly the source of the general rules of morality" (*TMS* VII.3.2.7). However, borrowing heavily from Francis Hutcheson's analysis of reasons for action in section 1

of his *Illustrations on the Moral Sense*, by which the latter was able to reduce practical rationality to its instrumental dimension, Smith insists on the inability of reason to provide, by itself, an end the desire of which would constitute an incentive to action: "But reason cannot render any particular object either agreeable or disagreeable to the mind for its own sake. Reason may shew that this object is the means of obtaining some other which is naturally either pleasing or displeasing, and in this manner may render it either agreeable or disagreeable, for the sake of something else" (*TMS* VII.3.2.7).

The source of motivation lies elsewhere than in reason, namely in the desire for what is pleasant and the aversion from what is painful. Consequently, the empirical observations on feelings and desires summarized in the general maxims also reveal to political scientists and actors not only the motivations but also the evaluations that govern human behavior. Smith subscribes to the thesis of the practical impotence of reason, which is common to both Hume and Hutcheson, and, like them, he understands it as an inability of reason alone, without borrowing from feelings and desires, to apprehend values and motivate action. Neither is reason per se capable of evaluation, although it is capable of generalization:

> But though reason is undoubtedly the source of the general rules of morality, and of all the moral judgments which we form by means of them, it is altogether absurd and unintelligible to suppose that the first perceptions of right and wrong can be derived from reason, even in those particular cases upon the experience of which the general rules are formed. These first perceptions, as well as all other experiments upon which any general rules are founded, cannot be the object of reason, but of immediate sense and feeling. (*TMS* VII.3.2.7)

A major consequence for our purpose is that general rules of morality, however useful, cannot by themselves, unaided by some affective or conative support, provide a practical remedy for self-deceit.

Eric Schliesser (2017, p. 93) claims that following rules is "the means toward correction of such self-deception." This seems true, but how does such a miracle work in detail? Again, the case of the man of furious resentment illustrates, from a temporal perspective, how this affective and habitual support of the rules operates. If he finally achieves command of his passion, it is because he has learned from the past that indulging in it will render him blameworthy in the future. Interestingly, Smith seems to consider that the efficacy of general rules relies on the fact that they heighten a temporal conflict that one form of self-deceit—that in which the spectator's gaze is tainted by the natural view—tends to weaken:

> He cannot throw off altogether the awe and respect with which he has been accustomed to regard [the rule]. At the very time of acting, at the moment in which passion mounts the highest, he hesitates and trembles at the thought of what he is about to do: he is secretly conscious to himself that he is breaking through those measures of conduct which, in all his cool hours, he had resolved never to infringe, which he had never seen infringed by others without the highest disapprobation, and of which the infringement, his own mind forebodes, must soon render him the object of the same disagreeable sentiments. (*TMS* III.4.12)

More formally, this means that, at date t_0 (when the passion occurs), furious resenters prefer x to y but, thanks to the general rules of morality acquired previously, have a foreboding—though an imperfect one—that they will prefer y to x at date t_1 and that choosing x at t_0 will produce unavoidable painful regrets at t_1. These anticipated regrets

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

allowed by the rules seem to prompt furious resenters to abstain from revenge and to remain time-consistent.

Reconstructing Smith's Response to Traditional Objections: The Mixed View

No sooner do we think we have found a solution to the insufficiency of the ideal impartial spectator than we discover to our chagrin that the apparent solution is also affected by the same threat, as Smith points out:

> The man who acts according to the rules of perfect prudence, of strict justice, and of proper benevolence, may be said to be perfectly virtuous. But the most perfect knowledge of those rules will not alone enable him to act in this manner: his own passions are very apt to mislead him; sometimes to drive him and sometimes to seduce him to violate all the rules which he himself, in all his sober and cool hours, approves of. The most perfect knowledge, if it is not supported by the most perfect self-command, will not always enable him to do his duty. (*TMS* VI.3.1)

While knowledge of the general rules of morality makes it possible to anticipate the regret that will follow their violation, to Smith this does not seem sufficient to guarantee their respect and thus time consistency. This seems all the more problematic since, if exemplary individuals who embody all the virtues except that of self-command are vulnerable to time inconsistency, then how much more true must this be for the rest of us, the weak "bulk of mankind" whose morality depends entirely on the sense of duty.

The use of "general maxims" thus faces a family of traditional objections that Smith could not fail to have had in mind, and which are still relevant to any serious reflection on the authority of personal rules: the Butlerian objection, the Hobbesian objection, and the Hutchesonian objection.

All these objections suggest the need for a mixed theory of general rules in which the influence of rules on distant future conduct strongly depends on a psychological basis involving habits, specific emotions or affective dispositions, and beliefs. Let us deal successively with them.

(i) The Butlerian objection is that the general character of rules, rather than discouraging, may encourage the tendency to self-deception that is rooted in the natural partiality of human beings. Indeed, a hallmark of partiality is the tendency to approve in general of a rule of conduct that one does not adopt in particular. As Joseph Butler put it: "Hence arises in men a disregard of reproof and instruction, rules of conduct and moral discipline, which occasionally come in their way: a disregard, I say, of these, not in every respect, but in this single one, namely, as what may be of service to them in particular towards mending their own hearts and tempers, and making them better men" (Butler 2017, p. 85).

Individuals who believe and assert the general maxim may fail to include themselves among the persons concerned by it. The typical case, which is the occasion for Butler's sermon, is provided by the parable of the ewe lamb set out by the prophet Nathan to King David, in II Samuel 12,[10] in which David very strongly condemns a type of conduct that

[10] The prophet Nathan tells King David the story of the owner of a large flock who steals his only ewe lamb from a poor man. This story made David indignant at the injustice. Nathan awakens David from self-deception by pointing out that he should be indignant about his own behavior, since he has ravished Bathsheba, the wife of Uriah.

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

he does not perceive to be his own. So, what would be the point of using a general rule "to correct the too partial views which self-love might otherwise suggest" (*TMS* III.4.11, quoted above), if adhering to a general rule is not invulnerable to the same bias?

It seems to us that Smith is aware of this difficulty and that this is one of the reasons why he insists that the acceptance of general rules must be implemented by habits and training. When Smith writes that "those general rules of conduct, *when they have been fixed in our mind by habitual reflection*, are of great use in correcting the misrepresentations of self-love concerning what is fit and proper to be done in our particular situation" (*TMS* III.4.12; emphasis ours), the *when* clause should be construed as indicating a necessary condition. This is also the reason why the effective authority of moral maxims requires affective incentives beyond the mere understanding of the rule of action, as we will see with Smith's response to the Hutchesonian objection.

(ii) The Hobbesian objection is that promises to oneself are not binding, and more generally that personal rules as mere resolutions create no obligation to comply. As Hobbes put it: "Nor is it possible for any person to be bound to himself, because he that can bind can release; and therefore, he that is bound to himself only is not bound" (Hobbes 1991, p. 184).

Smith is not really concerned by this objection since he conceives general rules as social rules from the outset. We have linked them to personal rules because they are maxims of conduct that an individual adopts. But they are not personal in the sense that they are adopted in isolation and arbitrarily. On the contrary, since social experience is the object of common knowledge, the rules that summarize it are both personal and social. The case of politeness is quite paradigmatic in this respect. Everyone can rely on these rules to avoid trouble or embarrassment in the future, but their obligation is not created by personal adoption, since they are social rules. It is the commitment that is personal, not the content of the rule. An interesting consequence is that, unlike the failure to comply with one's purely personal and voluntary resolutions, the failure to observe a general maxim does not make future failures more probable, as the maxim remains in the repertoire of social patterns of conduct. Such a conclusion contradicts the game theory approaches adopted by Benabou and Tirole (2004) to represent a self-control problem based on internal commitment, or by Meardon and Ortmann (1996) to interpret the acquisition of self-command in *The Theory of Moral Sentiments*. Both rely on the influence of self-reputation or self-confidence on the respect of rules—understood as "personal rules"—so that any failure is likely to create a precedent.

(iii) The Hutchesonian objection is the most important, since it focuses on the core problem of motivation. It points out the fact that mere knowledge of a proposition, including of general maxims, has no motivational power per se. As Hutcheson put it: the knowledge of "relations of inanimate objects to rational agents" "neither excites to actions, nor justifies them, without presupposing either affections or a moral sense" (Hutcheson 2002, p. 158).

We believe that Smith pondered upon this difficulty and that it led him to account for the influence of general rules by reference to something other than a supposed intrinsic authority (as a deontological moral theory would require). Indeed, his theory of motivation is sentimentalist down to its very roots, and his solution to this particular difficulty is again sentimentalist. Thus, the key to his solution is to realize that here we are dealing with specific feelings and emotions linked to a sense of duty. Smith's response to the objection is twofold. It relies on the consideration of habits and of what

we call *deontic emotions*, emotions that are relating to the respect or transgression of rules. In Smith's allusive account, they include both the rule-focused emotion of "reverence" and sanction-focused fear and hope. The stabilization of action according to a general rule of conduct is, first of all, a matter of habit, insofar as the rule requires a correction of feelings. Only habit can effect an enduring change in the tendency to feel a certain way in a certain circumstance. People may be "accustomed to correct and regulate their natural sentiments by general rules" (*TMS* II.2.3.10). Habit, here, is not only a disposition to act in a certain way but also a disposition to have such feelings or emotions, and it is in this sense that even general rules of politeness, not only the "sacred rules of justice," can inspire "habitual reverence" (*TMS* III.5.2).

The person who is devoid of feelings of gratitude nevertheless feels obliged to express gratitude to her benefactor. She does this by virtue of the conditional law that if one is in the situation where one has benefitted from the generosity of others, one must express gratitude to the benefactor. But what is the motivation for her action? It is not the general rule itself but a feeling associated with it. The motive "may be no other than reverence for the established rule of duty, a serious and earnest desire of acting, in every respect, according to the law of gratitude" (*TMS* III.5.1). This reverence for the rule is part of the deontic emotions. The desire to act according to the rule and the tendency to experience this type of deontic emotion are implanted in a habitual form through training and education: "There is scarce any man, however, who by discipline, education, and example, may not be so *impressed* with a regard to general rules, as to act upon almost every occasion with tolerable decency, and through the whole of his life to avoid any considerable degree of blame" (*TMS* III.5.1; our emphasis).

It is important to understand that here the avoidance of blame is not the end pursued, in a sense that the desire to avoid blame would provide the motivation for rule-compliant action; rather, blame avoidance is the secondary result of a desire to conform to a rule that is independent of the desire to avoid blame.

The deontic emotion of "reverence" has the general rule as its object. The general rule must embody a value that is the typical object of this emotion. This is how we should understand Smith's repeated use of the qualifier "sacred" as applied to rules of morality in general, including justice.[11] A rule that really deserves reverence is sacred, and, conversely, reverence is the affective attitude that is appropriate to what is sacred.

If Smith's readers have not paid enough attention to the role deontic emotions play in the enforcement of general rules, it may be because the theme is associated by Smith with the religious question. In this context, Smith does not hesitate to speak of the "double confidence" we should have in religious people (*TMS* III.5.13). Smith thus moves away from the irony of his friend Hume, in section 11 of *An Enquiry Concerning Human Understanding*, regarding the ambition of revealed as well as natural religion to present

[11] Contrary to Fricke (2011), we do not consider that "sacred" points to the absolute authority that would be possessed only by rules of justice, as opposed to other, less demanding, rules of morality. Smith devotes an entire chapter to examining, case by case, the extent to which each virtue "arises chiefly or entirely" "from a regard to general rules" (*TMS* III.6.1). The sacred character of rules is not only that of the rules of justice (*TMS* III.5.1–2). Although "the most sacred regard" is due to the rules of justice (*TMS* III.6.10), the duties of gratitude are "perhaps the most sacred of all those which the beneficent virtues prescribe to us" (*TMS* III.6.9). It remains true that, for Smith, there is an intimate connection between a conscientious, exact application of the rule and sacredness. Sacredness is a matter of degree, depending on whether the rule requires more or less precise application.

itself as a condition, or at least as a support, of public morality. He could have emphasized certain deontic emotions only in relation to the prospect of a transgression of duty—one thinks, of course, of the feeling of guilt—but he chose to take his examples from the fear and hope of sanctions in a supposed future state, thus conforming to the orthodox expectations with which academics, in his time more than today, had to deal. However, as we finally show, the theoretical significance of the reference to the "belief of a future state" (*TMS* III.5.10) is accounted for in our construal of the sense of duty.

IV. THE SACRALIZATION OF RULES

We have confronted the problem posed by the fact, outlined by Smith, that the authority of the impartial spectator is insufficient to ensure the morality and the effective consistency of action over time. We have shown that Smith's solution of appealing to the general rules of conduct suffers, at first sight, from the same defect, as Joseph Butler had predicted. We conclude that Smith was aware of this difficulty and therefore insisted on the role of what we call "deontic emotions," without which general rules of morality would have little "influence" on conduct.[12] While general rules are formed as summaries of socially shared evaluative experience, and are in this sense oriented towards the past of interactions and evaluations, once endorsed as practical maxims they require the conformity of the future to the past and thus demand consistency over time. We draw attention to the fact that Smith explicitly relates the constancy of choices not directly to general rules but to a specific attitude towards them, a "sacred regard":

> Without this sacred regard to general rules, there is no man whose conduct can be much depended upon. It is this which constitutes the most essential difference between a man of principle and honour and a worthless fellow. The one adheres, on all occasions, steadily and resolutely to his maxims, and preserves, through the whole of his life, one even tenor of conduct. The other acts variously and accidentally, as humour, inclination, or interest, chance to be uppermost. (*TMS* III.5.2)

Our analysis of this attitude as a set of habitual beliefs about duty combined with the disposition to experience deontic emotions not only dismisses the superficial view that the virtue of self-command is the final word in Smith's theory and the solution to all difficulties, including the practical difficulty of temporal inconsistency, but also qualifies the privilege and centrality that is—not without reason—accorded to the spectatorial theory. It also tempers an interpretation that draws the elaborations on general rules in *The Theory of Moral Sentiments* closer to a Kantian ethics of duty, or even to Emmanuel Kant's views on respect for the moral law, such as that offered, albeit with some qualification, by Samuel Fleischacker (1991).

[12] Here we take the term in its motivational sense. Smith does not clearly distinguish between the two terms he uses, "authority" and "influence." "Authority" clearly refers to a normative grip (that of guiding power, showing the direction to be followed in our actions), but in some passages it is undistinguishable from a motivational pull (*TMS* VI.2.1.9). "Influence" seems to be primarily motivational: that which tends to modify passion or action effectively. But it is very difficult to distinguish between the two when Smith speaks of the "influence and authority of the general rules of morality" (*TMS* III.5), since, as we have argued, these rules are per se devoid of motivational force.

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

One might object that we should have drawn a distinction between the different types of general rules that Smith discusses: rules of prudence, beneficence, and justice. We are aware of the distinctive position of the rules of justice in his work. However, Smith treats them jointly with the other rules: most of the illustrations he gives when introducing the general rules of morality concern the rules of justice. They are one species of general rules and are based on the same epistemology as the rules of prudence and beneficence. The fact that they have a specific character—they are "exact and precise"—does not make them an exception: on the contrary. They are an extreme case that serves to illustrate the power that Smith attributes to deontic emotions in the mobilization of the will: "The most sacred regard is due to them," says Smith, "and the actions which this virtue [that of justice] requires are never so properly performed as when the chief motive for performing them is a reverential and religious regard for those general rules which require them" (*TMS* III.6.10).

This discussion leads us to pay attention to the importance, for Smith, of the social conditions of agency. When he speaks of self-command as a virtue—that is, not simply as the varying degree to which ordinary individuals, more or less subject to "the usual weakness of human nature" (*TMS* I.3.1.13), control their action over time—he is describing a human perfection that is rarely met in commercial societies. The happy few possessed of this virtue have the habit of organizing their conduct and attitudes from the perspective and through the authority of the ideal spectator. In contrast, if we understand self-command not as a superlative virtue but as a relative quality, a degree of self-control, it is fortunate that the rest of us may achieve it through means other than a Stoic power over one's passions. This is where general rules come in. But they cannot operate alone.

For we are extremely far from a situation in which fully autonomous agents voluntarily give themselves a rule of action. While few are truly capable of the virtue of self-command, which remains a moral ideal embodying a form of "manhood" ("the manhood of self-command," *TMS* III.3.34), all rely on rules without having the capacity to directly control their passions. It is at this point that the deontic emotions come into play, both the "reverence" for the rule and the emotions related to the expected or actual consequences of transgression or compliance with the rule. Although Smith reminds us in the classic terms of Enlightenment rhetoric of the dangers of religious superstition (*TMS* II.2.3.12), he gives an auxiliary role to the emotions traditionally associated with the anticipation of positive or negative sanctions linked to the observance or transgression of "sacred" rules, in line with his interest in religion, as Jeng-Guo Chen (2017, p. 58) puts it, "as a subjective need, not as an objective truth"[13]

To understand how the use of rules rests on very different principles from those of a spectatorial theory, it is important to note that the anticipation of these sanctions is independent of the aversion to blame or the desire for praise. The latter are desires relative to spectators' responses. The hope for or fear of sanctions associated with compliance or transgression is the anticipation of happiness or unhappiness, not of failure or success in meeting social expectations. Moreover, these attitudes toward the sanctions associated with the observance of rules are subservient to the reverence for

[13] In doing so, Smith rejects Shaftesbury's aristocratic indignation at any "mercenary" conception of morality (Gill 2020), and overcomes Hume's extreme rejection of any positive link between morality and religious passions (Russell 2008).

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

rules. However, hope for or fear of sanctions is not a de jure necessary condition of the sense of duty. The rule may well be an object of habitual reverence in itself. It is just that the common person de facto needs an additional source of motivation.

It may seem astonishing that a sentimentalist should have recourse to some of the advantages of divine command theory. This astonishment diminishes if we note that this recourse is limited to the motivational aspect, under a general condition of "usual weakness," for agents who do not have superlative self-command. Sacredness provides a new source of motivation that is completely independent of the correctness of the rules and may even foster their perversion: "That principle which gives the greatest authority to the rules of duty, is alone capable of distorting our ideas of them in any considerable degree" (*TMS* III.6.12). One might also ask why a conception that recognizes the moral importance of the ideal of autonomy (Chen 2017) gives a role to the prospect of divine sanctions and whether this is not inconsistent.[14] However, one could apply here the idea that the "theonomous" moral attitude is not necessarily heteronomous, provided that believing in the "divine sponsorship" of moral principles does not prevent one from valuing them for their own sake (Adams 1979, p. 194, drawing on Paul Tillich).

In this investigation, we have kept an eye on theoretical developments that, since the end of the twentieth century, have sought to establish the role of certain emotions in supporting personal resolutions in intertemporal choice and in intra- or interpersonal diachronic control. Whether concerning the various intrapsychic means of commitment (Ainslie 1992), certain affective dispositions such as the sense of guilt (Frank 1988), or personal rules (Bénabou and Tirole 2004), there is an important body of work, often at the edges of behavioral economics, which shows that Thomas Schelling set exaggerated limits upon the spectrum of effective means of choice control when he emphasized constraints and extrapsychic means. It seems to us that these avenues of research do not represent a break with the program opened up by *The Theory of Moral Sentiments* and that Smith's reflection on the value of "general maxims" remains instructive, especially when it leads to emphasis on the fact that the influence of rules on conduct is mediated by certain emotions, in accordance with the traditional, but still relevant, view that the effective means of resisting an affective temptation are of the same, affective, nature.

The temporality of action raises the possibility of inconsistency between important preferences and actual choices, which would constitute a more enigmatic case of irrationality if it were synchronic. Smith understood that the atemporal evaluation of the impartial spectator, however severe, does not provide a solution, except in the counterfactual case of eminently self-controlling agents, those who embody the virtue of self-command. Setting aside external commitment devices, the effective remedy for everyone lies in the use of general rules.

According to our mixed view, there is a set of necessary conditions for the effectiveness of such rules. General rules that are successful in ensuring constancy and consistency are not merely voluntary personal commitments but social norms endorsed by individuals, drawn from the common experience of real impartial spectators and a history of moral interactions, made habitual, and accompanied by future-oriented

[14] Chen writes: "The man within … helps us fulfill the duty of humans to act autonomously and righteously. In this context, religious sentiments are generated not from the lack of due rewards in this life, nor from fear of due punishment for their unfulfilment" (2017, p. 59). We do not think that the value of autonomy is undermined by the appeal to "sacred rules."

https://doi.org/10.1017/S1053837225100746 Published online by Cambridge University Press

emotional dispositions that motivate compliance. It is in this context that the reference to beliefs about divine sanctions should be understood, and we thus assign it theoretical significance beyond its rhetorical dimension: "This reverence is still further enhanced by an opinion which is first impressed by nature, and afterwards confirmed by reasoning and philosophy, that those important rules of morality are the commands and laws of the Deity, who will finally reward the obedient, and punish the transgressors of their duty" (*TMS* III.5.3).

The implementation of the rule, either through habitual reverence and an earnest desire to do one's duty, or more commonly through the anticipation of sanctions, is conditioned by beliefs about the authority of rules. When one has the belief that a moral rule is God-given, one tends to anticipate those sanctions as highly probable.

Two clarifications are necessary: First, when Smith speaks of belief in sanctions in a future state, he does not mean that we rely on divine justice to punish those who have wronged us. On the contrary, he means that we avoid wronging others because we respect duty as sacred, and that this respect is reinforced by our possible belief that we will be subject to sanction in a future state. Second, although Smith refers to divine sanctions in a future state, he does not mean to say that the anticipation of sanctions associated with transgression of divine commandments is the only motivation for our acceptance of general rules and, in particular, rules of justice. It is a possible and fairly common motivation, but it is only one of the forms that respect for the rules takes. The basic mechanism seems to be more like this: sacralization is the representation of certain rules as divine, which gives rise to deontic emotions that favor the acceptance of these rules. These emotions include fear and hope but also simple reverence for duty. The claim we have been arguing for is also "mixed" in another respect: namely, that we have introduced a subcategory of subjective commitment devices, not objectively external but not purely internal either: rules represented as transcendent. Usually, when people form a resolution, they are aware of its internal character and have no trouble recognizing its being purely internal. However, in our analysis, rules can acquire strong practical authority when accompanied by the "impressed" belief that they are external to the agent and even to the "sentiments of mankind." Although these rules are self-imposed and can therefore be waived by the person concerned, they are perceived as being imposed by an authority that severely sanctions their infringement. It is important to emphasize, however, that belief in the divine origin of rules does not so much affect the degree of sanction expected as the probability with which one expects to be punished. The divine character bestowed on the rules does not necessarily lead to the anticipation of stronger sanctions; rather, it leads to the certainty of their occurrence because of the supposed omniscient and omnipotent nature of the divine. Drawing on Smith's vocabulary of the sacred, we may describe the process involving this type of habitual belief and the forward-looking deontic emotions it triggers as "sacralization." Sacralized rules are embodied by "important rules of morality" that are believed to be divine. The "sacred regard" is so deeply engrained, involving beliefs and affective dispositions that are quite beyond control, that it acts like a long-lasting determinant of the agent's choices.

We are now better able to understand where we started. It is not the maxim per se but "reverence for the rule," "impressed" first by nature and then by "past experience," that "checks the impetuosity" of passion (*TMS* III. 4.12, quoted above). Although rules of justice and morality are sacred, there is nothing magical about their authority.

COMPETING INTERESTS

The author declares no competing interests exist.

REFERENCES

Adams, Robert. M. 1979. "Autonomy and Theological Ethics." *Religious Studies* 15 (2): 191–194.

Ainslie, George. 1992. *Picoeconomics. The Strategic Interaction of Successive Motivational States Within the Person*. Cambridge: Cambridge University Press.

Ainslie, George, and Nick Haslam. 1992. "Self-Control." In Jon Elster and George Loewenstein, eds., *Choice Over Time*. New York: Russell Sage Foundation, pp. 177–209.

Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112 (4): 848–886.

Bianchini, Victor. 2016. "James Mill on Intemperance and Individual Preferences." *Journal of the History of Economic Thought* 38 (1): 21–40.

Bréban, Laurie. 2014. "Smith on Happiness: Towards a Gravitational Theory." *European Journal of the History of Economic Thought* 21 (3): 359–391.

Butler, Josep. 2017. *Fifteen Sermons Preached at the Rolls Chapel and Other Writings on Ethics*. Edited by David McNaughton. Oxford: Oxford University Press.

Chen, Jeng-Guo S. 2017. "'The Man Within': Adam Smith on Moral Autonomy and Religious Sentiments." *Journal of Scottish Philosophy* 15 (1): 47–64.

Elster, Jon. 1984. *Ulysses and the Sirens*. Revised edition. Cambridge: Cambridge University Press.

Fishburn, Peter C., and Ariel Rubinstein. 1982. "Time Preference." *International Economic Review* 23 (3): 677–694.

Fleischacker, Samuel. 1991. "Philosophy in Moral Practice: Kant and Adam Smith." *Kant-Studien* 82 (3): 249–269.

Frank, Robert. 1988. *Passions within Reason: The Strategic Role of the Emotions*. New York: Norton.

Fricke, Christel. 2011. "Adam Smith and 'the Most Sacred Rules of Justice.'" *Adam Smith Review* 6: 46–74.

Gill, Michael B. 2020. "Shaftesbury on Selfishness and Partisanship." *Social Philosophy and Policy* 37 (1): 55–79.

Hobbes, Thomas. 1991. *Leviathan*. Edited by Richard Tuck. Cambridge: Cambridge University Press.

Hutcheson, Francis. 2002. *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*. Edited by Aaron Garrett. Indianapolis: Liberty Fund.

Jaffro, Laurent. 2022. "Weakness and the Memory of Resolutions." In Carla Bagnoli, ed., *Time in Action. The Temporal Structure of Rational Agency and Practical Thought*. New York: Routledge, pp. 221–242.

———. 2024. *Le* Miroir *de* la *Sympathie. Adam Smith et le Sentimentalisme*. Paris: Vrin.

Konow, James. 2012. "Adam Smith and the Modern Science of Ethics." *Economics and Philosophy* 28 (3): 333–362.

Liu, Glory M., and Barry R. Weingast. 2020. "Deriving 'General Principles' in Adam Smith: The Ubiquity of Equilibrium and Comparative Statics Analysis Throughout His Works." *Adam Smith Review* 12: 134–165.

Meardon, Stephen. J, and Andreas Ortmann. 1996. "Self-Command in Adam Smith's Theory of Moral Sentiments. A Game-Theoretic Reinterpretation." *Rationality and Society* 81 (1): 57–80.

Montes, Leonidas. 2020 "The Relevance of Propriety and Self-Command in Adam Smith's Theory of Moral Sentiments." *Social Philosophy and Policy* 37 (1): 118–137.

Palacios-Huerta, Ignacio. 2003. "Time-Inconsistent Preferences in Adam Smith and David Hume." *History of Political Economy* 35 (2): 241–268.

Rawls, John. 1955. "Two Concepts of Rules." *Philosophical Review* 64 (1): 3–32.

Remow, Gabriela. 2007. "General Rules in the Moral Theories of Smith and Hume." *Journal of Scottish Philosophy* 5 (2): 119–134.

Russell, Paul. 2008. *The Riddle of Hume's Treatise: Skepticism, Naturalism and Irreligion*. New York: Oxford University Press.

Schelling, Thomas. 1984. "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." *American Economic Review* 74 (2): 1–11.

Schliesser, Eric. 2017. *Adam Smith: Systematic Philosopher and Public Thinker*. Oxford: Oxford University Press.

Smith, Adam. 1976. *The Theory of Moral Sentiments*. Edited by David D. Raphael and Alec L. Macfie. Oxford: Clarendon Press.