



HAL
open science

Reciprocity in Smith

Benoît Walraevens

► **To cite this version:**

Benoît Walraevens. Reciprocity in Smith. *Œconomia - History/Methodology/Philosophy*, 2020, 10 (4), pp.657-686. 10.4000/oeconomia.9771 . hal-03127781

HAL Id: hal-03127781

<https://hal.science/hal-03127781>

Submitted on 1 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Les contenus d'*Economia* sont mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Reciprocity in Smith

Benoît Walraevens*

This paper aims at providing a reconstruction of a general theory of reciprocity from Smith's works which provides both a descriptive and a normative account of reciprocity, and explains reciprocal behavior in bilateral relationships as well as the origin and evolution of social rules of reciprocity. I will show that this "Smithian" theory of reciprocity can explain several experimental results, generate previsions for new experiments and offer an interesting alternative to existing models of reciprocity and social preferences in economics. Smith offers a non-utilitarian, emotion-based and moral theory of reciprocity. We conclude in underlining how the idea of reciprocity is at the heart of Smith's theory of morality and of his moral egalitarianism.

Keywords: Smith (Adam), reciprocity, morality, social preferences, experiment

La réciprocité chez Smith

Cet article a pour but de reconstruire une théorie générale de la réciprocité à partir des textes de Smith. Cette théorie présente à la fois une approche descriptive et normative de la réciprocité, et explique aussi bien les comportements de réciprocité dans les relations entre deux individus que l'origine et l'évolution de règles sociales de réciprocité. Nous montrerons que cette théorie « smithienne » de la réciprocité peut expliquer de nombreux résultats expérimentaux, générer des prévisions pour de nouvelles expériences et offrir une alternative aux modèles existants en économie, les modèles de « préférences sociales ». Smith offre donc une théorie non utilitariste et morale de la réciprocité, fondée sur les émotions de gratitude et de ressentiment. Nous montrons enfin que l'idée de réciprocité est au cœur de la théorie morale de Smith et de son égalitarisme moral.

Mots-clés : Smith (Adam), réciprocité, moralité, préférences sociales, expérimentation

JEL: B12, B31

*Assistant Professor in Economics, University of Caen-Normandie, CREM, UMR CNRS 6211, Caen, France. benoit.walraevens@unicaen.fr

I would like to thank Jean Dellemotte and two referees of this journal for helpful comments on previous versions of this paper. All remaining errors are mine.

While reciprocity has long been an object of study in both the natural and the social sciences,¹ economics has been late to study it.² But since the 1990s, a significant number of empirical and theoretical works have been published on this topic, providing the economists with different models of reciprocity. These recent developments are part of their growing interest in non-selfish or “other-regarding” preferences and motivations. Economists had to explain why many people in laboratory and field experiments often do not behave according to the predictions of the *homo oeconomicus* model and, especially, why they reward co-operators and punish uncooperative people in so-called social dilemmas, even when they can expect no benefit from their behaviour in the future (*i.e.* in unrepeated games). Reciprocity naturally became a possible answer (among others) to this puzzle.

The purpose of this paper is twofold. Firstly, it aims at providing a reconstruction of a *general* theory of reciprocity from Smith’s works. By “general” I mean a theory which can provide both a descriptive and a normative account of reciprocity, and explain reciprocal behavior in bilateral relationships as well as the origin and evolution of social rules of reciprocity. Reciprocity is simply defined here as rewarding people who have acted kindly (“positive” reciprocity) and punishing those who have acted hurtfully (“negative” reciprocity), even though it can be costly for the reciprocator. As we shall see, while reciprocity can be seen as a form of exchange between individuals, it differs from “economic” exchange in important aspects and especially as in reciprocity people are treated as ends in themselves³ and not as means to satisfy one’s own ends (Kolm, 2000, 1). Even if Smith never used the words “reciprocity” or “reciprocal” in his books,⁴ I believe that the concept of reciprocity plays a major role in his explanation of social and moral life. It is a key concept for understanding the importance of the fundamental equality of individuals in his moral theory. Moreover, following Hume’s project to provide a new science of human nature based on empirical grounds⁵, Smith’s interest in reciprocity is to be seen as the expression of his careful study of the variety of people’s motivations for action, and especially of their other-regarding principles.

¹ See for example the pioneering works of Hamilton (1964) and Trivers (1971) in biology, Mauss (1924), Malinowski (1932) and Sahlins (1972) in anthropology, and Becker (1956) and Gouldner (1960) in sociology. For early works on reciprocity in the field of psychology, see Macaulay and Berkowitz (1970).

² A notable exception in economics is Kolm, whose works on reciprocity date back to the 1970s and early 1980s and whose 2008 book might be the most comprehensive study on reciprocity in economics.

³ This view of reciprocity, in which other people are ends for us rather than means, is called “intrinsic reciprocity” by Sobel (2005) and contrasted with selfish or “extrinsic reciprocity” in which other people reciprocate because it is in their interest to do so.

⁴ However, he did use the term “reciprocally” once (*TMS*, II.iii.3.1, 85).

⁵ On Smith’s attempt at building his own science of human nature, see Berry (2014).

I will show that this “Smithian” theory of reciprocity can explain several experimental results,⁶ generate previsions for new experiments and offer an interesting alternative to existing models of reciprocity and social preferences in economics, as Smith and Wilson (2017; 2019) recently showed. Indeed, Smith offers a non-utilitarian, emotion-based and moral theory of reciprocity, as we will see later. This paper thus completes Paganelli (2011)’s more general study of how Smith’s *Theory of Moral Sentiments* (*TMS*) can explain human cooperation in both personal and impersonal exchanges. Secondly, I will highlight that the idea of reciprocity as such is at the heart of Smith’s moral theory, our moral judgments on others and on ourselves crucially depending on our constant switch of position between agent and spectator.

I will build my account of Smith’s model of reciprocity mainly, though not only, upon his analysis of the merit and demerit of actions as it is presented in his *TMS*. Smith’s moral treatise is now widely quoted by economists, and especially by behavioral and experimental economists (Khalil, 1990; 1996; 2017; Ashraf, Camerer, and Loewenstein, 2005; Gintis *et al.*, 2005; Sally, 2001; Smith, 1998; 2003; 2010; 2016; Smith and Wilson, 2014; 2017; 2019), who find important insights therein regarding the psychology and morality of human decision-making. More specifically, I will focus my attention on the two passions that are the cornerstones of positive and negative reciprocity: gratitude and resentment, respectively. In Smith’s taxonomy of the passions, gratitude belongs to the “social” passions which bind people together, whereas resentment is one of the “unsocial” passions which oppose us. This point leads to important differences between positive and negative reciprocity, as we will show later. Yet the common characteristic and specificity of these passions lies in the fact that in addition to elicit our interest in others’ welfare, they push us to be the instrument of it, that is, to make them feel joy or pain for what they have done to us or to other people we sympathise with (*TMS*, II.ii.1.5, 68). Gratitude and resentment have received extensive but generally distinct treatments in Smithian scholarship,⁷ except in Elster (2011), Smith (2016), and Smith & Wilson (2017; 2019). The present study intends to provide a more detailed and comprehensive view of Smith’s conception of reciprocity, and especially of its moral character. I will thus complete and sometimes criticize the interpretations of these authors in

⁶ Smith’s works can also explain several experimental results related to the consequences of markets on morality, as was shown by Paganelli (2013).

⁷ On gratitude in Smith, see Hanley (2009) and Harpham (2004; 2012), who considers that Smith might be the greatest modern philosopher of gratitude (2004, 28) and that gratitude is at the heart of his moral theory (2004, 22). On the latter point, we will try to show that it is reciprocity in general, and not just gratitude, which is the core of Smith’s moral theory. Concerning Smith’s theory of resentment and its antecedents, specific studies include Pritchard (2008), Stalley (2012), MacLachlan (2010), and Schwarze and Scott (2015).

order to create a general and singular theory of reciprocity from Smith's texts. The paper is organised as follows. Section 1 and 2 will present Smith's descriptive approach to positive and negative reciprocity between two individuals. Section 3 discusses indirect reciprocity, or what happens when a third individual, an external, real, and uninvolved spectator is introduced into the model. Section 4 then details Smith's normative point of view on reciprocity, or what he defines as a proper reciprocity between two individuals. This leads to an introduction of Smith's concept of the impartial spectator's perspective. Section 5 enlarges the model to an indefinite number of individuals, explaining the foundations and evolution of social norms of reciprocity. Section 6 sums up the main characteristics of Smith's general model of reciprocity in modern terms and underlines how it substantially differs from existing models in economics. Moreover, I show how this model provides a wide range of testable propositions and explains a large class of experimental results. I conclude in underlining how the idea of reciprocity is at the heart of Smith's theory of morality and of his moral egalitarianism. Judgments of propriety and the figure of the impartial spectator both rely upon a constant and reciprocal switch of position between actor and spectator.

1. Positive Reciprocity Between Two Agents: a Descriptive Approach

I begin my analysis of Smith's theory of reciprocity with his description of the simplest, paradigmatic case of reciprocity, namely positive reciprocity between two agents. It is based on the natural feeling of gratitude.⁸ It takes place when an agent performs a kind action towards another agent.⁹ Let us call them respectively *A* and *B*. The latter, whose wellbeing has been enhanced by *A*'s generous action, will feel some gratitude towards *A*. Consequently, (s)he will desire to reward *A* and thus will reciprocate their kind action.

Let us focus first on *A*'s behaviour—the benefactor. The reciprocity relationship is opened by *A*'s generous or kind action towards *B*. Smith makes two interesting claims about benefactors. Firstly, that “Nature ... exhorts mankind to acts of beneficence, by the pleasing consciousness of deserved reward” (*TMS*, II.ii.3.4, 86). In other words, the pleasure which is derived from being beneficent takes two distinct forms: the first form lies in a pleasure of self-approbation not unlike what economists call a “warm-glow” effect (Andreoni, 1990) and the second lies in a pleasure in being approved by others for our kind action. The

⁸ Smith's views on gratitude seem especially indebted to the philosophers of ancient Rome and Greece, and especially Seneca and Cicero (Elster, 2011; Harpham, 2012; Tegos, 2018).

⁹ On Smith's theory of the virtue of beneficence, see Hanley (2009, chapter 6).

latter is a consequence of sympathy, allowing us to share others' passions and feelings. The former comes from the fact that *A* knows that his or her conduct is approved of by the impartial spectator (*TMS*, II.ii.2.4, 85), so that "when he looks forward to those whom he has served, [he] feels himself to be the natural object of their love and gratitude, and, by sympathy with them, of the esteem and approbation of all mankind" (*ibid.*). In Smith's words, (s)he knows (s)he is "praiseworthy". In fact, properly beneficent actions generally attract two kinds of non-monetary rewards: social praise and praiseworthiness, which Smith takes pains to distinguish and to hierarchize (*TMS*, III.2.1-2, 113-114). Indeed, praiseworthiness is, for Smith, of higher value than and can be enjoyed independently from social praise (*TMS*, III.2.5, 115-116). Moreover, he adds that the pleasure of social approbation is of little value when it is not coupled with the pleasure of self-approbation (*TMS*, III.2.4, 114-115). We need to think that we deserve praise to enjoy others' praise fully. Smith's second claim about benefactors has to do with the end of this quote and helps to make a connection with reciprocators' behaviour. Indeed, as Kolm (2008, 78-79, 140) also noticed, Smith states that the value of counter-gifts by direct and indirect reciprocators will significantly exceed the cost of our initial gift.¹⁰ But if there is a net benefit in giving, does that mean that for Smith, giving could be self-interested? Of course not. The importance of motives, which we will insist on later, must be introduced here. If we give in order to be given something in return, or in other words if we exhibit selfish motives, we will not be rewarded as we should be, because gratitude would be diminished or non-existent. We conclude that "selfish" giving would be ineffectual and self-defeating.

This leads us to focus on reciprocators' behaviour. Reciprocity seems to be a natural and universal phenomenon for Smith, as it is for Hutcheson or Hume. He writes that "Nature, which formed men for that mutual kindness, so necessary for their happiness, renders every man the peculiar object of kindness, to the persons to whom he himself has been kind." (*TMS*, VI.ii.1.19, 225) In other words, "kindness is the parent of kindness" (*ibid.*) and thus we are naturally generous towards those who have been generous with us, and with others (*TMS*, II.ii.1.10, 82). As we shall see, Smith's *TMS* abounds with such maxims of reciprocity¹¹. The analysis of positive reciprocity is based on the passion of gratitude, which prompts us to "reward" our (properly motivated,

¹⁰ "No benevolent man ever lost altogether the fruits of his benevolence. If he does not always gather them from the persons from whom he ought to have gathered them, he seldom fails to gather them, and with a tenfold increase, from other people." (*TMS*, VI.ii.1.19, 225)

¹¹ For example, Smith quotes the Christian command "to love our neighbour as we love ourselves" but adds that "it is the great precept of nature to love ourselves only as we love our neighbour, or what comes to the same thing, as our neighbour is capable of loving us." (*TMS*, I.i.5.5, 25)

unselfish) benefactors.¹² But what exactly makes us feel grateful? Everything that causes us some pleasure (*HA*, III.2, 49) and emanating from “sensitive” beings, Smith answers,¹³ because the satisfaction of gratitude precisely consists in “retaliating those sensations upon what gave occasion to them; which it is to no purpose to attempt upon what has no sensibility.” (*TMS*, II.iii.1.3, 94) Receiving a gift creates a feeling of “debt”, of which we will not be unloaded until we have “recompensed” our benefactor or “repay[ed]” him for his generous conduct (*TMS*, II.i.1.5, 68). In the case of multiple benefactors, Smith notes, men tend to think that their gratitude and “debt” is “divided among the different persons who contributed to their pleasure” and thus that “a smaller share seems due to anyone.” (*TMS*, II.iii.2.2, 98)

Yet Smith goes further, arguing that returning the pleasure we had is neither the sole nor the true end of positive reciprocity. In Smith’s words, “what gratitude chiefly desires, is not only to make the benefactor feel pleasure in his turn, but to make him conscious that he meets with this reward on account of his past conduct, to make him pleased with that conduct, and to satisfy him that the person upon whom he bestowed his good offices was not unworthy of them.” (*TMS*, II.iii.1.4, 95) Reciprocity is a very specific type of exchange, a *moral* and *personal* relationship in which we share our common humanity and attain human dignity as persons of equal worth. As Smith writes, we are happy to find someone whose sentiments are in agreement with our own, who values us as we value ourselves “and [who] distinguishes us from the rest of mankind” (*ibid.*). Our benefactor makes us feel loved and esteemed and it is to “maintain in him these agreeable and flattering sentiments” and to show that we were not “unworthy of them” that we return his favours (*ibid.*). Altruism being a scarce resource (*TMS*, III.3.4, 137), we admire the capacity of the beneficent man to transcend his natural preference for himself, his self-love (Hanley, 2009, 158). Reciprocating gifts and benefits fosters social concord and harmony.

However, there are exceptions to this natural gratitude and reciprocity towards our benefactors. Sometimes people are not rewarded for their kindness. They face “ingratitude”, which is strongly disapproved of by the impartial spectator owing to its improper, selfish motives (*TMS*, II.ii.1.3, 78; VI.ii.1.19, 225). This means that not returning presents exposes individuals to social disapprobation and blameworthiness. We reciprocate other people’s generosity also to avoid the pain

¹² “To reward, is to recompense, to remunerate, to return good for good received.” (*TMS*, II.i.1.4, 68) Hanley (2009, 156, 180) also underlines that Smith’s analysis of gratitude is founded on the idea of reciprocity but we disagree with him when he suggests that reciprocal beneficence is, for Smith, self-interested or based on mutual advantage (2009, 181).

¹³ “Before anything can be the proper object of gratitude or resentment, it must not only be the cause of pleasure or pain, it must likewise be capable of feeling them.” (*TMS*, II.iii.1.3, 94)

of feeling guilty. Even though of “all the duties of beneficence, those which gratitude recommends to us approach nearest to what is called a perfect and complete obligation”, we cannot compel someone to recompense us for our beneficence (*TMS*, II.ii.1.3, 79). The want of gratitude “cannot be punished” because it “does no positive hurt to anybody” (*TMS*, II.ii.1.3, 79). It just disappoints our expectations. Beneficence “is always free, it cannot be exerted by force” (*ibid.*, 78). Moreover, it would be improper to ask someone to reward us for our generosity and even “more improper than his neglecting to perform it” because it would mean that we made a gift to get a counter-gift, which we characterised as “selfish” giving (*TMS*, II.ii.1.3, 79).

2. Negative Reciprocity Between Two Agents: a Descriptive Approach

Smith’s descriptive point of view on negative reciprocity parallels his account of positive reciprocity. Just as people feel some natural gratitude towards their benefactors, they will feel some natural resentment towards individuals who hurt them.¹⁴ Smith claims that “As every man doth, so shall it be done to him, and retaliation seems to be the great law which is dictated to us by Nature” (*TMS*, II.ii.1.10, 82). This maxim of reciprocity implies here that if *A* harms *B*, *B*’s resentment towards *A* will lead him to punish *A*. For Smith “resentment seems to have been given us by nature for defence, and for defence only” (*TMS*, II.ii.1.4, 79). Despite being an unsocial passion, resentment thus plays a seminal role in the preservation of social order and that is why it is so important for Smith.

It is nothing but “the safeguard of justice” because “it prompts us to beat off the mischief which is attempted to be done to us, and to retaliate that which is already done” (*ibid.*). Resentment for Smith is both backward *and* forward looking¹⁵, making the offender “repent of his injustice” and preventing others from being tempted to harm us as well “through fear of like punishment” (*ibid.*).¹⁶ Mirroring the analysis

¹⁴ Smith’s conception of resentment seems to be highly indebted to Butler’s. For a comparison between Butler and Smith on resentment, see Pritchard (2008) and MacLachlan (2010).

¹⁵ Stalley (2012, 76) also notes that for Smith, resentment is not only ‘backward-looking’.

¹⁶ Interestingly, Smith defends this view of negative reciprocity in the *WN* on a critical issue of economics: international commerce. He takes the example of the commercial war between France and Great Britain in the seventeenth and eighteenth centuries, accusing the French of imposing high duties or prohibitions on the import of British products. He argues: “Revenge in this case naturally dictates retaliation, and that we should impose the like duties and prohibitions upon the importation of some or all of their manufactures into ours.” (*WN*, IV.ii.38, 467). That is precisely what happened, leading to mounting hostility between these countries. For Smith, “There may be good policy in retaliations of this kind when there is a

of gratitude, resentment is excited by everything that causes us some pain and that originates from sensitive beings (*TMS*, II.iii.1.1-2, 94). Moreover, its satisfaction will also consist in “retaliating those sensations upon what gave occasion to them”, that is in this case to cause the person who hurts us to feel pain (*TMS*, II.iii.1.3, 94). Also, the more pain we will have endured, the greater our natural resentment and subsequent punishment will be (*TMS*, II.ii.2.2, 83). However, once again, returning the sensations we have had does not seem to be the true goal of (negative) reciprocity for Smith (*TMS*, II.iii.1.5, 95). Smith provides us with a distinctive theory of punishment (Stalley, 2012, 69). Beyond inflicting some pain, punishing our offenders for their bad behaviour is meant to make them repent of their conduct (*TMS*, II.iii.1.5, 96). In particular, Smith argues, they must understand that no one is entitled to value themselves more than others and to satisfy their self-love regardless of the consequences of their actions on others. The issue of human equality and dignity is here at the forefront of reciprocity¹⁷. People who hurt us must be aware that all individuals are of equal worth and thus deserve equal respect. As was rightly noted by MacLachlan, resentment contains “a desire for accountability and acknowledgment from the wrongdoer that she be made to grieve on account of her behaviour toward me” (MacLachlan, 2010, 165). Reciprocity is an ethical relationship because it involves our seeing other people as ends, rather than as means for satisfying our own interests. In Smith’s words, “What chiefly enrages us against the man who injures or insults us, is the little account which he seems to make of us ... the unreasonable preference which he gives to himself above us, and that absurd self-love, by which he seems to imagine, that other people may be sacrificed at any time, to his conveniency or his humour.” (*TMS*, II.iii.1.5, 96) The offender did not put himself into others’ position and so did not take into account the consequences of his conduct upon them. While the benefactor exhibited *less* self-love than what we naturally expect from our fellows, here the offender shows *excessive* self-love and an unforgivable contempt for the happiness of others. He “wants *that sense of what is due to our fellow-creatures* which is the basis of justice and of society” (*TMS*, II.iii.2.8, 102, we underline). Revenge and punishment are precisely meant to “bring him back to a more just

probability that they will procure the repeal of the high duties or prohibitions complained of”, that is, only if we can expect these retaliations to lead to mutual cooperation and free commerce in the future (*WN*, IV.ii.39, 468). This will depend on the “skill” of “that insidious and crafty animal, vulgarly called a statesman or legislator, whose councils are dictated by the momentary fluctuations of affairs.” (*WN*, IV.ii.39, 468)

¹⁷ The most extensive treatment of Smith’s conception of dignity is to be found in Debes (2012) who argues that its main feature is to be based on the affective dimension of man rather than the rational or divine part of man in the kantian and theological traditions.

sense of *what is due to other people*, to make him sensible of *what he owes us*, and of the wrong that he has done to us" (TMS, II.iii.1.5, 96, we underline). "A moral being is an accountable being", as Smith nicely writes elsewhere (TMS, 111). Our natural capacity of resentment presupposes a sense of common humanity (Schliesser, 2017, 90). There are things that we owe to other people, just because they are our fellows, and especially not to hurt them for improper reasons. With reciprocity we try to restore the natural equality of men and to make them feel that they are "but one of the multitude, in no respect better than any other" (TMS, III.3.4, 137). He must understand "the deformity of doing the smallest injury to another, in order to obtain the greatest benefit" for himself (*ibid.*). Resentment is a call for equal, mutual respect and reciprocal recognition (Darwall, 2006, 84-85). We feel some resentment because we get less than what is our due, while we feel gratitude because we get more. Our rewards and punishments restore our natural equality.

Besides, an essential common feature of positive and negative reciprocity, of the natural feelings of gratitude and resentment, is their dependence upon both the intentionality of and motivations behind the actions that benefited or harmed us. As Smith clearly states: "before anything, therefore, can be the complete and proper object, either of gratitude or resentment", it must have produced pleasure or pain "from design" and, he adds, "from a design that is approved of in the one case, and disapproved in the other" (TMS, II.iii.1.6, 96). We can draw several consequences or propositions from this point. Firstly, it means that if we think that we deserved to be harmed, we should not feel some resentment towards our offender (TMS, II.iii.1.5, 96). Secondly, that if we judge the motivations of our benefactor to be proper, and that of our aggressor to be improper and yet the person "failed in producing either the good or the evil which he intended", then "less gratitude seems due to him in the one [case], and less resentment in the other" (TMS, II.iii.1.6, 96). Thirdly, that if we disapprove of the motives of our benefactor and do not disapprove of those of our offender and yet their actions produced "either great good or great evil", then we will feel "some gratitude" in the former case and "some resentment" in the latter (*ibid.*, 96-97). A fourth proposition, unmentioned here by Smith but discussed later, is that unintentional though beneficial or hurtful actions will elicit less gratitude and resentment.¹⁸ We will come

¹⁸ As Schliesser (2017, 121-133) nicely shows, Smith offers an interesting analysis of the peculiar feeling, distinct from guilt, which is what we feel when we have involuntarily hurt someone (TMS, II.iii.3.4-5, 106-107; VII.iv.30, 338-339). It leads to a feeling of debt and to a willingness to help the victim or his brethren. Consequently, it gives rise to good offices, like gratitude does. Schliesser (2017, 133) and Hankins (2016, 739) rightly note that this sentiment is quite important for developing our sense of human dignity and equal respect for others because through it man is "taught to reverence the happiness of his brethren, to tremble lest he should,

back to the issues raised by intentionality and motivations when we explain Smith's normative point of view on reciprocity between two individuals. Before that, we need to understand "indirect" reciprocity, that is, reciprocity by external observers who are not directly benefited or harmed.

3. Indirect Reciprocity

Therefore, indirect reciprocity is a refinement of our model, introducing new agents, external spectators who reciprocate kind and hurtful actions done to others. Meaning that if *A* helps *B*, then *C* (and *D*, *E*, etc.) will reward and help *A*. Likewise, if *A* hurts *B*, then *C* (and *D*, *E*, etc.) will punish and hurt *A*. In his book on reciprocity, Kolm (2008, 5) calls this phenomenon "reverse reciprocity" or the "Descartes effect" because it was first noticed by the French philosopher in his work on the passions of the soul. Moreover, we concur with Kolm (2008, 78) that Smith also underlined this form of indirect reciprocity in almost the same terms. There are therefore always three points of view on any given situation: the viewpoint of the agent, that of the affected party, and that of unaffected spectator(s), the external observers (Harpham, 2012, 14-15). Two issues will be discussed here. Firstly, why is it that unaffected, external spectators will exert some reciprocity? Secondly, what are the consequences of indirect reciprocity? As to the first point, the causes of indirect reciprocity, Smith's explanation is based on his analysis of individuals' natural sharing of passions: the faculty of sympathy or "fellow-feeling". In the case of benevolent actions, an external, uninvolved spectator will feel what Smith calls a "redoubled sympathy" (*TMS*, I.ii.4.1, 38). This means that he will approve of and sympathise with both the kindness and generosity of the benefactor and with the gratitude of the beneficiary (*ibid.*, 39; *TMS*, II.i.2.4, 70). The necessary condition for the natural gratitude of the spectator to emerge and for reciprocity to take place is that the individual approves of the motives behind the generous act of *A* towards *B* (*TMS*, II.i.3.1, 71). It is pleasant for a spectator to observe benevolent actions and the pleasure (and gratitude) it gives to the beneficiary. In the case of malevolent actions, an external, uninvolved spectator will feel a "divided sympathy" (*TMS*, I.ii.3.1, 34). This means that (s)he will disapprove of and feel some antipathy towards the offender and will sympathise with the resentment of the victim (*ibid.*). As Smith states, people have a strong "fellow-feeling with the injuries that are done to their brethren" (*TMS*,

even unknowingly, do any thing that can hurt them, and to dread that animal resentment which, he feels, is ready to burst out against him, if he should, without design, be the unhappy instrument of their calamity" (*TMS*, II.iii.3.4, 106). Yet we agree with Hankins (2016, 73), against Schliesser (2017), that for Smith the pious feeling has more to do with regret than with shame.

I.iii.2, 34). This is “indignation”, or third-party resentment (*TMS*, I.ii.3.3, 35). Again, the assessment of the motives behind the action that elicited some resentment and demand for reciprocity is crucial. In other words, external observers will feel some indignation only insofar as they judge the intentions of the offender to be improper (*TMS*, II.i.3.1, 71; II.i.4.3, 74). Otherwise, they would deserve no punishment, as we saw earlier (*TMS*, II.i.3.3, 72). Another common point between the gratitude and resentment (or indignation) of the spectator is that both are necessarily less intense than the “original” passion of the agent, since they are a “copy” of it in the former’s imagination.

There is, nonetheless, an important difference between indirect positive and negative reciprocity: it is much easier to sympathise with gratitude than with resentment, so that indirect gratitude will be more frequently encountered than indignation. The reason is that resentment is part of the unsocial passions, whose propriety point is, as Smith claims, very low. It thus follows that we need to impose serious limits on the intensity of our (natural) resentment if we want others to share it (*TMS*, I.ii.3.1, 34). By contrast, “generosity, humanity, kindness ... all the social and benevolent affections, when expressed in the countenance or behaviour ... please the indifferent spectator upon almost every occasion.” (*TMS*, I.ii.4.1, 39) Moreover, even when they are “excessive”, the benevolent passions “are never regarded with aversion” (*TMS*, I.ii.4.3, 40). On the contrary, even though “mankind ... have a strong sense of the injuries that are done to another” (*TMS*, I.ii.3.2, 34), excessive expressions of “anger” and “resentment” are clearly frequent and condemned (*TMS*, I.ii.3.4, 35).¹⁹

Concerning my second point, the consequences of indirect reciprocity, I have already partly answered this question in the case of positive reciprocity in section I. Here, I underlined that for Smith there is a net benefit in giving because the value of the counter-gifts by several reciprocators significantly exceeds the cost of the initial gift, whether or not they directly benefited from it (*TMS*, VI.ii.1.19, 225). External spectators, who easily sympathise with beneficiaries’ gratitude, will thus take an active part in reciprocity, even though their indirect gratitude is less intense than the original one. We can call this phenomenon “third-party reward”.²⁰ Consequently, there is a pleasure for the benefactor in receiving approval not only from the recipient of their kind act, but also from these indirect reciprocators (additional and distinct from the benefactor’s pleasure of self-approbation). Along the same lines, in the case of negative reciprocity the spectators’ indignation will prompt them to ask for punishment and/or punish the offender

¹⁹ Likewise, “a person becomes contemptible who tamely sits still, and submits to insults, without attempting either to repel or to revenge them. We cannot enter into his indifference and insensibility: we call his behaviour mean-spiritedness” (*TMS*, I.ii.3.3, 35).

²⁰ On this indirect generosity towards strangers, see Ule et al. (2009).

themselves.²¹ This is what is now called “third-party punishment”. However, knowing that this indirect resentment is necessarily less intense than the original resentment of the victim, third-party punishment will be less severe than second-party punishment. The same reasoning applies to third-party reward, which will be less intense than the reciprocity of the beneficiary. Besides, someone who has performed malevolent acts will, at a minimum, be disapproved of and ostracized by these external, indifferent spectators. Even if this individual is not actually punished for these acts, they will be haunted by their conscience because they recognise themselves as blameworthy. Thus indirect reciprocity creates significant reasons for dissuading people to hurt others on the one hand, and for encouraging their benevolence on the other.

4. The Moral Sentiments of Gratitude and Resentment: the Normative Account of Reciprocity

The first three sections focused on Smith’s descriptive analysis of the *natural* feelings of gratitude and resentment, which lead to direct and indirect, positive and negative reciprocity. However, as Elster (2011) also noticed, Smith additionally offers a normative point of view on reciprocity when he discusses the *moral* sentiments of gratitude²² and resentment and the proper reward and punishment for acts of kindness and malevolence. Unsurprisingly, the norm of reciprocity is founded on the impartial spectator’s point of view, which is the equivalent in our imagination of the external, uninvolved observer we studied in the previous section. So, in order to feel and express (perfectly) proper degrees of gratitude and resentment we need to imagine what an external, unaffected and yet well-informed spectator would feel and do in our case (*LJ(B)*, 475). This process involves a distancing from our natural station and feelings of gratitude and resentment. More precisely, we need to restrain the intensity of our passion so that every impartial spectator can sympathise with it (*TMS*, II.i.5.7, 76). The effort of self-command we are asked to apply here varies greatly in terms of the degree of passion we feel. Inspired by Aristotle, Smith defines the point of view of the impartial spectator as a kind of golden mean between the excess and lack of a passion. Too much resentment and retaliation

²¹ “When we see one man oppressed or injured by another, the sympathy which we feel with the distress of the sufferer seems to serve only to animate our fellow-feeling with his resentment against the offender. We are rejoiced to see him attack his adversary in his turn, and are eager and ready to assist him whenever he exerts himself for defence, or even for vengeance within a certain degree.” (*TMS*, II.i.2.5, 70-71)

²² Our distinction between “natural” and “moral” sentiments of gratitude is similar to Harpham’s (2012, 19) distinction between “unrefined” and “refined” sentiments of gratitude in Smith.

will be disapproved of and condemned, as will insufficient revenge. Knowing that the propriety point of resentment is quite low, as we saw in section 2, it means that we have to make huge efforts of self-command in order to significantly decrease the intensity of that passion if we wish to obtain the approbation of the impartial spectator. Expressing a proper or just degree of resentment and hence of negative reciprocity is thus much more difficult than doing the same for gratitude and positive reciprocity. This is due to the asymmetrical effect of pleasure and pain on people's well-being, as recent works confirmed (Offerman, 2002). For Smith, pains are more intensely felt than pleasures and thus "adversity ... necessarily depresses the mind of the sufferer much more below its natural state, than prosperity can elevate him above it" (*TMS*, I.iii.1.3, 44; I.iii.1.8, 45). As a conclusion, individuals often feel excessive resentment towards their offenders and might then exert undue revenge if there were no "judges" to prevent them from doing so. The character of the judge is essential here. This individual embodies the incarnation of the uninvolved, well-informed, and impartial spectator and as such can decide what the just punishment for our aggressor is (*TMS*, Appendix II, 389). Resentment is the "safeguard of justice", Smith claims (*TMS*, II.ii.1.4, 79). It is fundamental for the peace, order, and stability of society to have judges who can relate and impose the good and just thing to be done for repairing harmful actions. If people took their own revenge, their natural tendency to exert excessive resentment would make society chaotic (*TMS*, Appendix II, 389). By way of contrast, no positive harm is done when our gratitude is defective and an excess of it is unproblematic. Consequently, we do not need a judge to impose duties of gratitude or to condemn ingratitude. This is useless and practically impossible because the rules of beneficence and gratitude do not have the same precision as the rules of resentment and justice, as we will see again later. To sum things up, in negative reciprocity the agent will pay his "moral debt" towards the victim to compensate for the damage done, whereas in positive reciprocity the beneficiary will "repay" his benefactor, both cases restoring the moral equality of men. With reciprocity we give people what they morally deserve.

But what are the information and criteria upon which the impartial spectator can judge of the propriety and justice of rewards and punishments, and thus of reciprocity? The judgment of the impartial spectator on the merit and demerit of actions and thus on reciprocity is mainly based on the three elements that we mentioned earlier. First, it depends on the *consequences* of the action on our well-being, whether it caused us some pain or pleasure. And there a simple rule applies: the more pleasure (pain) we had, the more gratitude (resentment) we will feel for our benefactor (offender), and the greater our reward (punishment) will be. If the agent failed to help or hurt us, gratitude and resentment will strongly diminish. Secondly, impartial spectators need to assess

the (*im*)propriety of the motives of our benefactor (offender), to see whether they (dis)approve of his/her intentions in helping (hurting) us. If egoistic reasons are discovered behind the beneficence or gratitude of someone, such as giving in order to receive in return or to impose on others and show his/her superiority, only “little gratitude seems due” in such cases (*TMS*, III.5.1, 162; *TMS*, II.i.3.1, 71-72). Likewise, “a very small return seems due to that foolish and profuse generosity which confers the greatest benefits from the most trivial motives, and gives an estate to a man merely because his name and surname happen to be the same [as] those of the giver” (*TMS*, II.i.3.2, 72). Moreover, satisfying gratitude out of the “cold sense of duty” is only a second best for the proper and just gratitude of the impartial spectator (*TMS*, III.5.1, 162). Concerning negative reciprocity, if we have been hurt for proper motives, “all sort of resentment seems unjust”, and punishment in this case is undeserved (*TMS*, II.i.3.1, 71-72). Thirdly, identifying proper returns for acts of beneficence and mischief requires taking into account the *intentionality* of these actions. If we have been involuntary hurt or benefited, we will feel less gratitude and resentment because the lack of intentionality diminishes the merit or demerit of the action (*TMS*, II.iii.1.6, 96). Note that Smith is aware of the issue of information asymmetry between givers (takers) and receivers (victims). In an ideal world, we should judge of the merit and demerit of an action on intentions alone. But men’s motivations for action are not always easily discernible and, Smith adds, if we had to punish every person showing bad intentions, the courts of justice would soon be full (*TMS*, II.iii.3.2, 105)! That is why Nature, in its “unerring wisdom”, made us judge of the merit of an action also and mostly on its consequences, which are more visible and straightforward. And this “irregularity of sentiments” is socially useful because “man was made for action” and so cannot “be satisfied with indolent benevolence, nor fancy himself the friend of mankind, because in his heart he wishes well to the prosperity of the world.” (*TMS*, II.iii.3.3, 106)

In short, the degree of reciprocity one deserves depends upon and varies with the intentionality, motivation, and outcome of the action. Yet Smith adds a fourth, crucial element, or rather a rule to apply in order to identify what is a proper reward or punishment, namely equality. He writes: “as soon as we can, we should make a return of equal, and if possible of superior value to the services we have received, would seem to be a pretty plain rule, and one which admitted of scarce any exceptions” (*TMS*, III.6.9, 174). Here we have what seems to be a precise and quite common rule of positive reciprocity: returning at least an equal value. However, the natural question that immediately comes to mind is the following: what should be made equal? And for whom? We can imagine several forms of equal value. Let us suppose that *A* makes a gift to *B* and *B* wants to return it. As was rightly

noted by Elster (2009, 149-150), the general rule of equal value could mean that:

- i. the value of the gift for *A* should be equal to the value of the counter-gift for *B*
- ii. the value of the gift for *A* should be equal to the value of the counter-gift for *A*
- iii. the value of the gift for *B* should be equal to the value of the counter-gift for *B*
- iv. the value of the gift for *B* should be equal to the value of the counter-gift for *A*

Smith is well aware of the ambiguities and indeterminacy of this so-called rule of equal or higher value in reciprocity,²³ which distinguishes the latter from economic exchange. He uses different examples to illustrate his point, one of them being particularly noteworthy. Indeed, Smith asks, “if your friend lent you money in your distress, ought you to lend him money in his? How much ought you to lend him? When ought you to lend him? Now, or to-morrow, or next month? And for how long a time?” (*TMS*, III.6.9, 174). For Smith, it is also obvious that “no general rule can be laid down, by which a precise answer can, in all cases, be given to any of these questions” (*ibid.*). The impartial spectator always takes into account the specificities of each case. His recommendations are highly context-dependent, taking into account the “character” of the persons and their personal “circumstances”, so that “you may be perfectly grateful, and justly refuse to lend him a half-penny: and, on the contrary, you may be willing to lend ... him ten times the sum which he lent you, and yet justly be accused of the blackest ingratitude” (*ibid.*).

What about negative reciprocity? Smith refers once to the idea that there should be “an equality at least betwixt the sufferings of the injur’d person and the offender” but he discusses here the special case of “wilful murder” and it is unclear whether this principle can be generalised to every sort of crime or injury (*LJ(A)*, 106). Going further, Elster (2011) claims that Smith would be favourable to *lex talionis*, the law of retaliation, which prescribes to take “an eye for an eye”. We disagree with Elster on this point. Firstly, Smith never explicitly claimed in his works that the law of retaliation is the right punishment for every crime and offence. Secondly, even though Smith claimed that death penalty could be approved of by the impartial spectator, as Elster (2011) rightly noticed, this is so in the case of wilful murder and, again, it is far from clear whether this can be generalised (*LJ(A)*, 10). Given that intentionality and motivations must be taken into account for judging the propriety of reciprocity, we think that Smith would not see

²³ “Upon the most superficial examination, however, this rule will appear to be in the highest degree loose and inaccurate, and to admit of ten thousand exceptions.” (*TMS*, III.6.9, 174)

death penalty as a proper punishment for every murder, and especially for unpremeditated and/or unintentional murder. The only measure of the right punishment is and always will be how far the impartial spectator can sympathise with the resentment of the victim: “if the injury is so great as that the spectator can go along with the injured person in revenging himself by the death of the offender, this is the proper punishment, and what is to be exacted by the offended person or *the magistrate in his place who acts in the character of an impartial spectator*” (LJ(A), 104, we underline). But if the spectator “could go along with him if he revenged the injury by a small corporal punishment or a pecuniary fine, this is the punishment that ought to be inflicted” (*ibid.*). I thus conclude that Smith’s normative principle of reciprocity, based on the approbation of the impartial spectator, is not always aligned with *lex talionis*, especially when the offender had “proper” motivations to hurt someone.

5. The Origin and Evolution of Social Rules of Reciprocity

What happens now to reciprocity if we repeat the interactions and deal with new people? While the impartial spectator’s point of view is necessary to define what a perfectly proper reciprocity is in a specific case involving two individuals, taking into account the circumstances of action and the characteristics of the persons, this effort of distancing from our natural point of view is out of reach for most of us, Smith claims. Only the wise and the virtuous always follow their inner voice. Most people, by contrast, are often prone to self-deceit, this “fatal weakness of mankind”, which makes us unable to judge ourselves with impartiality (TMS, III.4.8, 158). Yet, Nature has offered us a “remedy” to our tendency to hide to ourselves our “deformities” of character: the “sense of duty”, by which we adhere to general rules of morality we have formed from “our continual observations upon the conduct of others” and thus can act with “tolerable decency” throughout our life (TMS, III.4.7, 158; III.5.1, 163). Some actions “call forth our approbation, and we hear everybody around us express the same favourable opinion concerning them” (*ibid.*). They want “to honour and reward them” because they excite the “love, the gratitude, the admiration of mankind”, so that “we become ambitious of performing the like; and thus naturally lay down to ourselves a rule ... that every opportunity of acting in this manner is carefully to be sought after” (*ibid.*). In other words, “we do not originally approve or condemn particular actions; because, upon examination, they appear to be agreeable or inconsistent with a certain general rule. The general rule, on the contrary, is formed, by finding from experience, that all actions of a certain kind, or circumstances in a certain manner, are approved or disapproved of” (TMS, III.4.8, 158). Smith thus presents a proto-evolutionary view of morality. The general rules of morality are conventions, emerging from the

“concurring sentiments of mankind” (TMS, III.4.11, 160). They are the unintended result of human social interactions, or more precisely of *n* players’ games of reciprocity and serve as a second best for people prone to “the misrepresentations of self-love” (TMS, III.4.12, 160). The latter memorize their experiences of approbation and disapprobation as both agent and spectator and infer general rules of conduct from them, which they strive to follow in order to maximize their chances of social approval (*i.e.* of receiving praise and praiseworthiness, and of avoiding blame and blameworthiness) without incurring the cost of the processing of information and careful analysis associated with the procedure of the impartial spectator, which deals with the specificities of each case.²⁴

As an illustrative example, Smith precisely discusses the “man of furious resentment” who, “if he was to listen to the dictates of that passion, would perhaps regard the death of his enemy as but a small compensation for the wrong, he imagines, he has received” (*ibid.*). However, “his observations upon the conduct of others, have taught him how horrible all such sanguinary revenges appear” and thus “he has laid it down to himself as an inviolable rule, to abstain from them upon all occasions” (*ibid.*). Thanks to his or her “reverence for the rule”, (s)he is made “incapable of being guilty of such a violence” (*ibid.*). This reasoning applies to positive reciprocity too. Sometimes, Smith writes, “the man who has received great benefits from another person, may, by the natural coldness of his temper, feel but a very small degree of the sentiment of gratitude” (TMS, III.5.1, 162) and “though his heart therefore is not warmed with any grateful affection, he will strive to act as if it was, and will endeavour to pay all those regards and attentions to his patron which the liveliest gratitude could suggest” (*ibid.*). Most importantly, “he will carefully embrace every opportunity of making a proper return for past services” and this “without any selfish intention of obtaining new favours” or “any design of imposing upon his benefactor or the public” (*ibid.*). The ultimate motivation of his action “may be no other than a reverence for the established rule of duty”, that is, to act according to “the law of gratitude” (*ibid.*). Thus, we think that Elster (2011) was wrong when he claimed that Smith never mentioned the idea that gratitude could be “enforced by social norms”. Once we have internalized them, social rules help us to mimic what others generally expect of us. But Smith makes it clear that this “sacred” respect for the rules of society, this “sense of duty”, as important as it is for the morality of society, is only a second best, especially in the case of positive reciprocity. In his words, “a benefactor thinks himself but ill requited, if the person upon whom he has bestowed his good

²⁴ For more details on the formation of social rules of cooperation in Smith, see Paganelli (2011).

offices, repays them merely from a cold sense of duty, and without any affection to his person" (*TMS*, III.6.4, 172).

Finally, Smith identifies a crucial difference between the rules of justice, of which resentment is the foundation, and the rules of social virtues like gratitude and benevolence. The former are like "the rules of grammar", they are "precise, accurate and indispensable" while the latter always remain "loose, vague and indeterminate, and present us rather with a general idea of the perfection we ought to aim at, than afford us any certain and infallible directions of acquiring it", like the rules of style or elegance in writing (*TMS*, III.6.11, 175-176). In other words, "there are no rules by the knowledge of which we can infallibly be taught to act upon all occasions with ... proper beneficence" (*ibid.*, 176). As outlined earlier, returning at least an equal value to our benefactor seems to be the most obvious of rules of positive reciprocity or gratitude and yet it allows for many exceptions and interpretations. By contrast, the rules of resentment and punishment are precise and can thus be codified, becoming the laws of justice, which are "the pillar" of society, whereas rules of beneficence are only its "ornament" (*TMS*, II.ii.3.4, 86). To conclude on this point, the general rules of positive and negative reciprocity, of resentment and gratitude, are nothing but the generalisation of the two players' interactions that we studied earlier. They are the necessary though unintended result of the convergent experiences of people in the indefinitely repeated game of social life.

6. Smith's Model and Reciprocity in Economics

Smith's theory of reciprocity presents three main characteristics that distinguish it from current theories of reciprocity in economics. The first characteristic, judiciously noted by Smith and Wilson (2014; 2017; 2019), is the fact that it is not based on the usual representation of individuals as utility-maximizers. It is a non-utilitarian theory of reciprocity and as such provides a fruitful alternative to social preferences models, most of which are based on utility maximization.²⁵ As we saw before, rendering the pain or pleasure we get from others' actions is part of reciprocity but it is not its ultimate foundation. Reciprocity takes into account not only the consequences or outcome of the action affecting us (and others), but also the intentionality and motivations behind it. By contrast, most economic models of reciprocity focus either on outcomes (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) or on perceived intentions (Rabin, 1993; Falk *et al.*, 2003; Dufwenberg and Kirchsteiger, 2004). However, one notable exception is Falk and Fischbacher's model (2006) which takes into account both intentions and outcomes. Interestingly, this model was then used and refined by

²⁵ For general criticism on social preferences models and their inability to adequately model reciprocity, see Wilson (2008; 2010).

Stanca *et al.* (2009) to include motivations and to create experiments testing the importance of the latter on reciprocity. Even though these models see humans as utility maximizers, and as such fundamentally differ from Smith's model, they provide some support for his general idea that reciprocal behavior relies on outcomes, intentionality, and motivations. In reciprocity, the *process* by which an outcome is reached is as important as the outcome itself. This is another difference with economic exchange and that is why the identification of people's intentions and motivations is so important for understanding reciprocal behaviour.

Moreover, while most economic models of reciprocity are descriptive and often built as post-hoc rationalizations of experimental data, Smith's theory is more than that. It is also prescriptive or normative, defining what a proper reciprocity is, and it goes deep into the causes and foundation of reciprocity. Finally yet importantly, for Smith, people will reciprocate actions even though it is costly to them, that is, without necessary regard for future benefits. Reciprocity is not (necessarily) for Smith a disguised selfish act, preserving our long-term interest. To use contemporary words, Smith's theory of reciprocity is based on "fairness" rather than on "mutual advantage". The second, notable characteristic of Smith's model of reciprocity is the fact that it is based on human emotional states or "feelings", which are generally ignored in economic theories of reciprocity.²⁶ As we have explained, the natural sentiments of gratitude, resentment, and indignation are the foundations of reward and punishment, and thus of reciprocity. People are driven by their passions to treat others as others treat them. The third characteristic of Smith's theory of reciprocity is that it is primarily a *moral* theory of reciprocity, meaning that reciprocal behaviour is not ultimately based on utilitarian considerations, but rather on the idea of the equal dignity of all human beings, who are moral creatures deeply interested in the propriety of their conduct.²⁷ Above all, they

²⁶ A notable exception is Cox *et al.* (2007), which provides a parameterised model of reciprocity in which emotional states like gratitude are included within utility functions.

²⁷ While they do not raise the issue of human equality and dignity in Smith, Smith and Wilson (2017; 2019) are perceptive, however, in their insistence on the importance of the propriety of conduct in Smith's theory of action. Along these lines, they offered a simple formal representation of Smith's original model of individual choice of action with its insistence on the propriety of conduct. It is presented as follows. An action a_i by individual i depends on i 's judgment of its propriety, given the action's contextual circumstances:

$$a_i(\text{Propriety} \mid C) = \alpha_i(C) \cdot PR + \beta_i(C) \cdot PR \cdot PW + \gamma_i(C) \cdot PW + \delta_i(C),$$

"where PR and PW are (0, 1) indicators that action is praised by others (1), or not (0), and is praiseworthy (1), or not (0); and $\alpha_i, \beta_i, \gamma_i$ and δ_i are non-negative functions weighting PR and PW in determining its propriety. In the second term, PW leverages PR , while the third expresses the sentiment that PW may yield stand-alone

want to be praised and praiseworthy and avoid being blamed and blameworthy. To sum up, Smith offers a non-utilitarian (and non-consequentialist), emotions-based, and moral theory of reciprocity, which has no equivalent in economics.²⁸ As we shall show now, Smith's theory of reciprocity can be empirically tested by means of traditional experiments like (modified) prisoner's dilemma games, trust games or ultimatum games. From our analysis of Smith's texts in the previous sections, we can now infer the following list of propositions on reciprocity:

- i. the sum of counter-gifts generally exceeds the value of the gift
- ii. in case of multiple benefactors, our gratitude and reciprocity is divided between them
- iii. properly motivated and voluntary beneficial actions deserve and elicit full gratitude and reward
- iv. improperly motivated, voluntary hurtful actions deserve and elicit full resentment and punishment
- v. properly motivated hurtful actions should not elicit any resentment
- vi. properly motivated yet unsuccessful benevolent actions elicit less gratitude than successful actions
- vii. improperly motivated yet unsuccessful harmful actions elicit less resentment and punishment than actually harmful actions
- viii. improperly motivated yet greatly beneficent actions elicit some gratitude
- ix. properly motivated and greatly hurtful actions elicit some resentment
- x. beneficial yet unintentional actions will elicit less gratitude and reward than intentional (and beneficial) ones
- xi. hurtful yet unintentional actions will give rise to less resentment and punishment than intentional actions
- xii. third-party reward is based on indirect gratitude and is less intense than second-party reward
- xiii. third-party punishment is based on indignation and is less intense than second-party punishment
- xiv. excessive resentment elicits no indignation and thus no third-party punishment
- xv. the value of the counter-gift should at least be equal to the value of the gift
- xvi. ingratitude cannot be punished or exerted by force because it makes no positive harm and beneficence is always free.

value, distinct from *PR*, even where no praise is possible." (Smith and Wilson, 2017, 35-36)

²⁸ For another singular and general theory of reciprocity, see Bruni (2008).

In joint work with Bart Wilson, Vernon Smith, whose interest in the works of Adam Smith is not new²⁹ but seems to have grown over the years, recently provided empirical tests and support for some of these propositions (Smith and Wilson, 2015; 2017; 2018; 2019).³⁰ For example, they tested Smith's idea that beneficence cannot be extorted by force with a modified ultimatum game (Smith and Wilson, 2018). Likewise, they imagined and implemented variations of trust games (Smith and Wilson, 2017; 2019) to test his parallel claims that "actions of a beneficent tendency, which proceed from proper motives, seem alone to require reward", whereas "actions of a hurtful tendency, which proceed from improper motives, seem alone to deserve punishment" (*TMS*, II.ii.1, 112). Interestingly, they show how Smith's ideas lead to a different understanding of a basic trust game by "players" (Smith and Wilson, 2019, 114, 148).

Let's take an example with a simple sequential prisoner's dilemma game to illustrate their argumentation. If both players choose to cooperate, they get an outcome of 3. If both choose to defect, they get 1. If one of them chooses to cooperate while the other defects, the former gets 4 and the latter gets 0. We use a traditional backward induction reasoning. Let's take first the position of player 2. He will imagine what it is fair to do, that is, what an impartial spectator would approve or disapprove. If player 1 chooses to cooperate, his action is properly motivated and can be perceived as kind and invoke gratitude because he forewent a sure outcome (of 1) and offered player 2 the possibility to improve his own outcome (get 3 or 4). So, player 2's choice to reciprocate cooperation would be praiseworthy because both are better off (3 instead of 1) and he renounced to a higher income (of 4). By contrast, choosing to defect would be disapproved of by the impartial spectator and by player 1 because the latter would finish with nothing after foregoing a sure outcome. Now let's take the position of player 1. He knows that "kindness is the parent of kindness" and so he can reasonably expect that if he chooses to cooperate, he will logically be reciprocated.

Along the lines of Smith and Wilson one could imagine investment games (in which a first mover can send a sum of money, which is multiplied, to a second mover, who can then resend part or nothing of the sum to the first mover) with variants including punishment options for first movers and unintentional choices by the latter, in order to test Smith's propositions on reciprocal kindness and ingratitude. We could also implement a two-rounds dictator game in which the roles would be inverted between the first and the second round, the dictator of the first round becoming the receiver of the second. Importantly, the two players shouldn't know from the beginning of the experiment that

²⁹ See Smith (1998).

³⁰ See also Smith (2013; 2016).

there would be a second round so that the “generosity” of the dictator in the first round cannot be motivated by the desire to be reciprocated. The amount sent by the dictator during this round could be considered as properly motivated beneficence, thus removing the ambiguity of Smith and Wilson’s trust games in which the seemingly beneficent move by the first player can be motivated by the expectation of higher gains for himself too. By way of consequence, the amount sent by the dictator in round 2 of our game could be considered as “reciprocal kindness” or beneficence, rather than “reciprocal cooperation” as can be found in trust games.³¹

Moreover, let me recall that for Smith ingratitude should not be punished while injustice should be. On the latter point, using “take” games (in which a player can take part of the endowment of another player who can then reply by taking back from the first mover) may be relevant to assess the contexts in which people feel some resentment and want to punish injustice.

More generally, in line with the predictions of Smith’s model of reciprocity, a great number of laboratory and field experiments on several classes of social dilemmas (such as trust games, gift-exchange games, investment games, ultimatum, and prisoner’s dilemma games) have shown the impact of perceived intentions and motivations, in addition to outcomes, on reciprocal behaviour (McCabe et al., 2003; Cox et al., 2006; Charness, 2004; Stanca et al., 2009; Lacour, 2011; Rand *et al.* 2015; Toussaert, 2017; Orhun, 2018). They have also provided empirical support for the importance of third-party reward and punishment, or indirect reciprocity, for fostering and sustaining cooperation among humans in social dilemmas.³² Lastly, experiments confirmed the lower intensity of third-party punishment with regard to second-party punishment (Fehr and Fischbacher, 2004).³³ However, I believe that there is still some work to be done by experimental economists to test Smith’s other propositions or predictions on reciprocity, especially their emotional foundations, and to understand how and why people try to conform to what others expect from them (and what they think they should expect from them) in order to receive praise and praiseworthiness and to avoid blame and blameworthiness. According to Smith, people naturally reward praiseworthy and punish blameworthy behaviors, even though it is costly to them and without expecting future benefits from their actions. Thus, I agree with Elster (2011) in claiming that Smith is a precursor of what is now called “strong reciprocity”³⁴ in

³¹ We are borrowing here Isoni and Sugden’s (2019) distinction.

³² On the similarities and differences between Smith’s impartial spectator and empirical spectators in modern social science, see Konow (2012).

³³ But note that recently Leibbrandt and Lopez-Perez (2012) found that third parties often punish as intensely as second parties and that most of the former do not punish in a more impartial way than the latter.

³⁴ For a more nuanced view on this issue, see Clark (2009).

opposition to self-interested, “weak” reciprocity (Guala, 2012). I think that his works could be a source of inspiration and insights for those who see humanity as a cooperative species and Man as a *Homo Reciprocans* (Fehr and Gächter, 1998; Bowles and Gintis, 2002; 2008; 2013; Gintis *et al.*, 2005; Dohmen *et al.*, 2009), while looking for an alternative to the models of social preferences built upon utility maximization, along the lines of Vernon Smith and Bart Wilson’s recent works. For Smith there are natural principles in human nature which make us reciprocate both kind and hurtful actions. His works can help economists better understand the emotional and moral determinants of human decision-making. Recent empirical works on and models of “guilt-aversion” (Charness and Dufwenberg, 2006; Battigali and Dufwenberg, 2007; Hauge, 2016; Khlmetzki, 2016; Baloufatas and Sutter, 2017; Bellemare *et al.*, 2017), or the idea that individuals try to conform to others’ expectations and moral standards to avoid the painful sensation of guilt, seem to be one step in that direction, since they confirm Smith’s claims about the importance of social and self-approbation in human conduct.³⁵

Conclusion: Reciprocity as the Foundation of Smith’s Theory of Morality

Adam Smith’s works offer a uniquely comprehensive, sophisticated, and testable theory of reciprocity, built upon his careful and penetrating observations of human behaviour. Along the lines of Hume, Smith’s moral philosophy is primarily a descriptive, social psychology with strong empirical foundations. That is why Smith’s *TMS* is still so effective at predicting and explaining contemporary human behaviour, and especially other-regarding principles of action. His way of doing research and of building theories has much in common with what behavioural and experimental economists do today, even though he provides us with an alternative to the utility-maximizing models of social preferences used for explaining reciprocity today. Smith’s model of reciprocity is founded on the passions of gratitude and resentment, defining what proper reciprocity is from the point of view of a well-informed, external, impartial spectator taking into account the consequences, intentionality, and motivations of action. In reciprocity, the process by which outcomes are reached is as important as the outcomes themselves. General rules of positive and negative reciprocity naturally emerge from the convergent sentiments of individuals as to what is fit and proper to be done in different circumstances and these rules are then internalised by social actors. Reciprocity is both an individual

³⁵ On the significant impact of social approbation and disapprobation on human conduct in social interactions, see Lopez-Perez and Vorsatz (2010) and Leibbrandt and Lopez-Perez (2014).

motivation and a social fact. It is an essential part of a peaceful and flourishing social life for Smith and a key ingredient of his moral theory. As Mauss later wrote, “reciprocity is the human rock on which societies are built” (*The Gift*, quoted in Kolm, 2008). For Smith, the laws of justice are the codification of the norms of the impartial spectator about negative reciprocity or punishment in a given society and they are the ultimate foundation of the peace and order of society, while relations of positive reciprocity and gratitude are what makes a society flourishing and happy (*TMS*, II.ii.3.2-3, 85-86). Besides, reciprocity is not limited to returning the pain or pleasure we get from others’ actions. More fundamentally, reciprocity is about seeing others as our equals, sharing our common humanity and respect for their dignity. But where does this feeling of equality comes from?

In order to explain it, I would like to underline that Smith’s moral theory itself is built upon an idea of reciprocity between the judgment of the actor and that of the spectator. The judgment of “propriety”, the adequacy or “suitableness” of a passion to its object, is built upon the process of mutual sympathy. This process is founded on both a reciprocal change of position and a reciprocal adjustment of our feelings to those of the person we try to identify with (*TMS*, I.i.2-4, 13-23). It works as follows. On the one hand, there is the agent or “the person principally concerned”, who feels the “original” passion. On the other hand, there is the spectator who feels, almost instantly, a “sympathetic” passion, which is a “copy”, though imperfect, of the original. Indeed, a significant element of the model is the gap between the intensity of the agent’s and spectator’s passion (*TMS*, I.i.4.7, 21): the former is necessarily higher than the latter, because sympathy is only an imaginary and imperfect mechanism for putting oneself into another’s place. An observer cannot feel stronger than the person principally concerned. The spectator will never be able to enter fully into the agent’s character or to know exactly the objects and causes of his passion. However, even if the spectator will never feel the passion of the agent with the same intensity, “harmony” and “concord” are not out of reach. What will prompt both the agent and the spectator to regulate the intensity of their passion and tend toward equilibrium is the common interest they have in the typically human desire for, and pleasure, in sympathy (*TMS*, I.i.2, 13-6). By contrast, the inability to reach a “concord” or agreement upon moral sentiments is painful. There is a double incentive to try to reach an agreement. Therefore, by his self-command, the agent will try to reduce the intensity of his passion so that the spectator sympathizes with him (*TMS*, I.i.4.6-7, 21-2), while reciprocally the spectator will increase his own intensity of passion by trying to enter into every circumstance causing the passion. By making mutual concessions, they fill the gap and reach an impartial equilibrium standpoint called the “propriety point”. Mutual sympathy, approbation and pleasure will then arise for both. In order to understand the rise of the

other people's sense of equality and dignity, of what is due to them just because they are our fellows, we need, again, to look at the figure of the impartial spectator which is the product of this constant change of position between the actor and the spectator, between our national station or point of view and those of others.³⁶ In other words, the creation in our imagination of the impartial spectator is the natural and "insensible" consequence of our experiences of successful and unsuccessful processes of mutual sympathy.³⁷ The desire of mutual sympathy insensibly and naturally urges us to go beyond ourselves, to compare our own vision with that of others and then to adopt a more general and impartial point of view which they can approve of. This indifferent, "uninvolved" and "impartial" spectator represents "the other people" coming from "far and near" (Sen, 2009) or, in other words, representing "mankind". Thus for Smith our "conscience arises out of the constant interplay of observation and experience, of empathy for others' feelings and for others' probable responses to one's own" (Clark, 1992, 191). And by seeing ourselves with the eyes of the impartial spectator, we understand that we are "but one of the multitude", that we are all equal and cannot with impunity harm others for satisfying our own interest, as we saw earlier. It is the impartial spectator (and its representation by magistrates) which will help us to "humble to arrogance of our self-love" and to be deeply concerned with others' happiness. From our constant and reciprocal change of position we learn our sense of equality and from this sense of equality we reward our benefactors and punish our offenders, that is, we reciprocate. We have come full circle.

³⁶ For a similar idea that this continual change of position between the actor and the spectator is the source of our sense of equality for Smith, see Debes (2012).

³⁷ "In order to produce this concord, as nature teaches the spectators to assume the circumstances of the person principally concerned, so she teaches this last in some measure to assume those of the spectators. As they are continually placing themselves in his situation, and thence conceiving emotions similar to what he feels; so he is as constantly placing himself in theirs, and thence conceiving some degree of that coolness about his own fortune, with which he is sensible that they will view it. As they are constantly considering what they themselves would feel, if they actually were the sufferers, so he is constantly led to imagine in what manner he would be affected if he was only one of the spectators of his own situation. As their sympathy makes them look at it in some measure with his eyes, so his sympathy makes him look at it, in some measure, with theirs, especially when in their presence, and acting under their observation: and, as the reflected passion which he thus conceives is much weaker than the original one, it necessarily abates the violence of what he felt before he came into their presence, before he began to recollect in what manner they would be affected by it, and to view his situation in this candid and impartial light." (*TMS*, I.i.4.8, 22)

References

- Andreoni, James. 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401): 464-477.
- Ashraf, Nava, Colin Camerer, and George Loewenstein. 2005. Adam Smith, Behavioral Economist. *Journal of Economic Perspectives*, 19(3): 131-145.
- Baloufatas, Loukas and Matthias Sutter. 2017. On the Nature of Guilt Aversion: Insights From a New Methodology in the Dictator Game. *Journal of Behavioral and Experimental Finance*, 13: 9-15.
- Battigali, Pierpaolo and Martin Dufwenberg. 2007. Guilt in Games. *American Economic Review Papers and Proceedings*, 97(2): 170-176.
- Becker, Howard. 1956. *Man in Reciprocity: Introductory Lectures on Culture, Society and Personality*, New York: Praeger.
- Bellemare, Charles, Alexandre Sebald, and Sigfried Suetens. 2017. A Note on Testing Guilt Aversion. *Games and Economic Behavior*, 102: 233-239.
- Berry, Christopher. 2012. Adam Smith's Science of Human Nature. *History of Political Economy*, 44(3), 471-492.
- Bolton, Gary and Axel Ockenfels. 2000. ERC—A Theory of Equity, Reciprocity and Competition. *American Economic Review*, 90(1): 166-193.
- Bowles, Samuel and Herbert Gintis. 2002. Behavioral Science: *Homo Reciprocans*. *Nature*, 415: 125-128.
- Bowles, Samuel and Herbert Gintis. 2008. L'idéal d'égalité appartient-il au passé ? *Homo reciprocans vs Homo oeconomicus*. *Revue du Mauss*, 31: 209-241.
- Bowles, Samuel and Herbert Gintis. 2013. *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton: Princeton University Press.
- Bruni, Luigino. 2008. *Reciprocity, Altruism and the Civil Society: in Praise of Heterogeneity*. New York and London: Routledge.
- Charness, Gary. 2004. Attribution and Reciprocity in a Simulated Labor Market: an Experimental Investigation. *Journal of Labour Economics*, 22(3): 665-688.
- Charness, Gary and Martin Dufwenberg. 2006. Promises and Partnership. *Econometrica*, 74(6): 1579-1601.
- Clark, Henry C. 1992. Conversation and Moderate Virtue in Adam Smith's TMS. *The Review of Politics*, 54(2).
- Clark, Henry C. 2009. Adam Smith and Neo-Darwinian Debate Over Sympathy, Strong Reciprocity and Reputation Effects. *Journal of Scottish Philosophy*, 7(1): 47-64.
- Cox, James C. and Cary A. Deck. 2006. Assigning Intentions When Actions Are Unobservable: the Impact of Trembling in the Trust Game. *Southern Economic Journal*, 73(2): 307-314.
- Cox, James C., Daniel Friedman. and Steven Gjerstad. 2007. A Tractable Model of Reciprocity and Fairness. *Games and Economic Behavior*, 59(1): 17-45.
- Darwall, Stephen. 2004. Equal Dignity in Adam Smith. *Adam Smith Review*, 1: 129-134.

- Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect and Accountability*. Cambridge, MA: Harvard University Press.
- Debes, Remy. 2012. Adam Smith on Dignity and Equality. *British Journal for the History of Philosophy*, 20(1): 109-140.
- Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde. 2009. Homo Reciprocans: Survey Evidence on Behavioural Outcomes. *The Economic Journal*, 119(536): 592-612.
- Dufwenberg, Martin and George Kirchteiger. 2004. A Theory of Sequential Reciprocity. *Games and Economic Behavior*, 47(2): 268-298.
- Elster, Jon. 2009. *Le désintéressement. Traité critique de l'homme économique I*. Paris: Seuil.
- Elster, Jon. 2011. Reciprocity in Seneca and Smith. *Adam Smith Review*, 6: 152-171.
- Falk, Armin and Urs Fischbacher. 2006. A Theory of Reciprocity. *Games and Economic Behavior*, 54: 293-315.
- Falk, Armin, Ernst Fehr and Urs Fischbacher. 2003. On the Nature of Fair Behavior. *Economic Inquiry*, 41(1): 20-26.
- Fehr, Ernst and Klaus Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3): 817-868.
- Fehr, Ernst and Simon Gächter. 1998. Reciprocity and Economics: the Economic Implications of *Homo Reciprocans*. *European Economic Review*, 42(3-5): 845-859.
- Fehr, Ernst and Urs Fischbacher. 2004. Third Party Punishment and Social Norms. *Evolution and Human Behavior*, 25(2): 63-87.
- Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr (eds). 2005. *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press.
- Gouldner, Alvin W. 1960. The Norm of Reciprocity: a Preliminary Statement. *American Sociological Review*, 25(2): 161-178.
- Guala, Francesco. 2012. Reciprocity: Weak or Strong? What Punishment Experiments Do (and Do Not) Demonstrate. *Behavioural and Brain Sciences*, 35(1): 1-15.
- Hamilton, William D. 1964. The Genetical Evolution of Social Behavior. *Journal of Theoretical Biology*, 7(1): 1-16.
- Hankins, Keith. 2016. Adam Smith's Intriguing Solution to the Problem of Moral Luck. *Ethics*, 126(3): 711-746.
- Hanley, Ryan P. 2009. *Adam Smith and the Character of Virtue*. Cambridge: Cambridge University Press.
- Harpham, Edward J. 2004. Gratitude in the History of Ideas. In Robert A. Emmons and Michael E. McCullough (eds), *The Psychology of Gratitude*. Oxford: Oxford University Press, 19-36.
- Harpham, Edward J. 2010. Adam Smith's Lost World of Gratitude. In Benjamin Ginsberg and Gwendolyn Mink (eds), *Political Science as Public Philosophy. Essays in Honor of J. Lowi*. New York: Norton and Company, 345-363.
- Hauge, Karen E. 2016. Generosity and Guilt: The Role of Beliefs and Moral Standards of Others. *Journal of Economic Psychology*, 54: 35-43.

- Isoni, Andrea and Robert Sugden. 2019. Reciprocity and the Paradox of Trust in Psychological Game Theory. *Journal of Economic Behavior and Organization*, 167: 219-227.
- Khalil, Elias. 1990. Beyond Self-Interest and Altruism: A Reconstruction of Adam Smith's Theory of Human Conduct. *Economics and Philosophy*, 6(2): 255-273.
- Khalil, Elias. 1996. Respect, Admiration, Aggrandizement: Adam Smith as Economic Psychologist. *Journal of Economic Psychology*, 17(5): 555-577.
- Khalil, Elias. 2017. Socialized View of Man vs. Rational Choice Theory: What Does Smith's Sympathy Have to Say? *Journal of Economic Behavior and Organization*, 143: 223-240.
- Khalmetski, Kiryl. 2016. Testing Guilt Aversion With an Exogenous Shift in Beliefs. *Games and Economic Behavior*, 97: 110-119.
- Kolm, Serge-Christophe. 1984. *La bonne économie : La réciprocité générale*. Paris: Presses Universitaires de France.
- Kolm, Serge-Christophe. 2000. Introduction to the Economics of Reciprocity, Giving and Altruism. In Louis-André Gérard-Varet, Serge-Christophe Kolm and Jean Mercier Yhtier (eds), *The Economics of Reciprocity, Giving and Altruism*. London: MacMillan, 1-44.
- Kolm, Serge-Christophe. 2008. *Reciprocity*. Cambridge: Cambridge University Press.
- Konow, James. 2012. Adam Smith and the Modern Science of Ethics. *Economics and Philosophy*, 28(3): 333-362.
- Lacour, Pierre. 2012. Intentions and Ethical Behavior in Trust Games. *Journal of Socio-Economics*, 41(1): 55-63.
- Leibbrandt, Andreas and Raul Lopez-Perez. 2012. An Exploration of Third and Second Party Punishment in Ten Simple Games. *Journal of Economic Behavior and Organization*, 84(3): 753-766.
- Leibbrandt, Andreas and Raul Lopez-Perez. 2014. Different Carrots and Different Sticks: Do We Reward and Punish Differently Than We Approve and Disapprove. *Theory and Decision*, 76(1): 95-11.
- Lopez-Perez, Raul and Marc Vorsatz. 2010. On Approval and Disapproval: Theory and Experiments. *Journal of Economic Psychology*, 31(4): 527-541.
- Macaulay, Jacqueline and Leonard Berkowitz. 1970. *Altruism and Helping Behavior*. New York: Academic Press.
- MacLachlan, Alice. 2010. Resentment and Moral Judgment in Smith and Butler. *Adam Smith Review*, 5: 161-177.
- Malinowski, Bronislaw. 1932. *Crime and Custom in Savage Society*. London: Paul, Trench, Trubner.
- Marshall, David. 1984. Adam Smith and the Theatricality of Sentiments. *Critical Inquiry*, 10(4): 592-613.
- Mauss, Marcel. 1923-24. Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques. *L'année sociologique*, seconde série, 1: 30-186.
- Offerman, Theo. 2002. Hurting Hurts More than Helping Helps. *European Economic Review*, 46(8): 1423-1437.

- Paganelli, Maria P. 2011. The Same Face of the two Smiths: Adam Smith and Vernon Smith. *Journal of Economic Behavior and Organization*, 78(3): 246-255.
- Paganelli, Maria P. 2013. Commercial Relations: from Adam Smith to Field Experiments. In Christopher J. Berry, Maria P. Paganelli and Craig Smith (eds), *Oxford Handbook of Adam Smith*. Oxford: Oxford University Press, 333-350.
- Pritchard, Michael S. 2008. Justice and Resentment in Hume, Reid and Smith. *Journal of Scottish Philosophy*, 6(1): 59-70.
- Rabin, Matthew. 1993. Incorporating Fairness Into Game Theory and Economics. *American Economic Review*, 83(5): 1281-1302.
- Rand, David, Drew Fudenberg, and Anna Dreber. 2015. It's the Thought that Counts: the Role of Intentions in Noisy Repeated Games. *Journal of Economic Behavior and Organization*, 116: 481-499.
- Sahlins, Marshall. 1972. *Stone Age Economics*. New York: De Gruyter.
- Sally, David. 2001. On Sympathy and Games. *Journal of Economic Behavior and Organization*, 44(1): 1-30.
- Schwarze, Michelle A. and John T. Scott. 2015. Spontaneous Disorder in Adam Smith's *Theory of Moral Sentiments*: Resentment, Injustice and the Appeal to Providence. *The Journal of Politics*, 77(2): 463-476.
- Sen, Amartya. 2009. *The Idea of Justice*, Harvard: The Belknap Press of Harvard.
- Smith, Adam. [1790] 1982. *The Theory of Moral Sentiments (TMS)*. D. D. Raphael and A. L. Macfie (eds). Indianapolis: Liberty Fund.
- Smith, Adam. [1762-1766] 1982. *Lectures on Jurisprudence (LJ)*. R. Meek, D. D. Raphael and L. Stein (eds). Indianapolis: Liberty Fund.
- Smith, Adam. [1795] 1982. *History of Astronomy (HA)*. In W. P. Whitman (ed.) *Essays on Philosophical Subjects*. Indianapolis: Liberty Fund.
- Smith, Vernon. 1998. The Two Faces of Adam Smith. *Southern Economic Journal*, 65(1): 1-19.
- Smith, Vernon. 2010. What Would Adam Smith Say? *Journal of Economic Behavior and Organization*, 73(1): 83-86.
- Smith, Vernon. 2013. Adam Smith: From Propriety and Sentiments to Property and Wealth. *Forum for Social Economics*, 42(4): 283-297.
- Smith, Vernon. 2016. Adam Smith and Experimental Economics: Sentiments to Wealth. In R. Hanley (ed.), *Adam Smith. His Life, Thought and Legacy*. Princeton: Princeton University Press, 262-280.
- Smith, Vernon and Bart Wilson. 2014. Fair and Impartial Spectators in Experimental Economic Behavior. *Review of Behavioral Economics*, 1(1-2): 1-26.
- Smith, Vernon and Bart Wilson. 2017. Sentiments, Conduct and Trust in the Laboratory. *Social Philosophy and Policy*, 34(1): 25-55.
- Smith, Vernon and Bart Wilson. 2018. Equilibrium Play in Voluntary Ultimatum Games: Benevolence Cannot Be Extorted. *Games and Economic Behavior*, 109: 452-464.
- Smith, Vernon and Bart Wilson. 2019. *Humanomics. Moral Sentiments and the Wealth of Nations for the Twenty-First Century*. Cambridge: Cambridge University Press.

- Sobel, Joel. 2005. Interdependent Preferences and Reciprocity. *Journal of Economic Literature*, 43(2): 392-436.
- Stalley, Richard. 2012. Adam Smith and the Theory of Punishment. *Journal of Scottish Philosophy*, 10(1), 69-89.
- Stanca, Luca, Luigino Bruni and Luca Corazzini. 2009. Testing Theories of Reciprocity: Do Motivations Matter. *Journal of Economic Behavior and Organization*, 71(2), 233-245.
- Tegos, Spiros 2018. Libéralité et moeurs des 'rangs moyens' chez Adam Smith. In Eleonore Le Jallé and Fiona McIntosh-Varjabedian (eds), *Liberal, Libéralité, Libéralisme. Histoire et enjeux philosophiques, culturels, et littéraires*. Paris: Honoré Champion, 147-166.
- Toussaert, Séverine. 2017. Intention-based Reciprocity and Signalling of Intentions. *Journal of Economic Behavior and Organization*, 137: 132-144.
- Trivers, Robert. 1971. The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, 46(1): 35-57.
- Ule, Alajz, Arthur Schram, Arno Riedl, and Timothy Cason. 2009. Indirect Punishment and Generosity towards Strangers. *Science*, 326: 1701-1704.
- Wilson, Bart. 2010. Social Preferences Aren't Preferences. *Journal of Economic Behavior and Organization*, 73(1): 77-82.
- Wilson, Bart. 2008. Language Games of Reciprocity. *Journal of Economic Behavior and Organization*, 68(2): 365-377.