Invisible-hand Explanations

Dayer-Tieffenbach, Emma

Invisible-hand Explanations

Emma Tieffenbach

Thèse de doctorat ès Lettres
Université de Genève
2011

Directeur de thèse:
Kevin Mulligan (Université de Genève)

Président du jury:
Pascal Engel (Université de Genève)

Jury:
Francis Cheneval (Universität Zurich)
Annabelle Lever (Université de Geneve)
Uskali Mäki (University of Helsinki)
Leo Zaibert (Union College, Schenectady, New York)

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

My deepest gratitude goes to Kevin Mulligan who supervised the dissertation. I have greatly benefited from his well-known sharpness, insightful suggestions, generous support and true indulgence.

I have had the pleasure of discussing The invisible hand with Markus Haller, Vivian Mizrahi and Victoria Tschumi. I am very thankful to all of them for their interest in my research, for their encouragement, and for pressing me to sharpen my arguments by being their harshest critics.

I also want to extend special thanks to Nir Eyal for helping me to organize my thoughts and time when self-confidence had left me.

During a three-year collaboration at the EPFL, Julien Deonna deeply influenced the way I write, discuss, and think about philosophy. I am very grateful to his exceptional teaching skills and friendship.

I am indebted and very thankful to Otto Bruun who offered many highly valuable comments on the whole manuscript. Otto also proofread the dissertation, cleansing it of its countless inaccuracies, ambiguities, and stylistic infelicities. The remaining scoriae, if any, are the unfortunate result of last-minute additions.

I want to express my gratitude to Monique Canto-Sperber, Pierre Demeulenaere, Vincent Descombes, Nir Eyal, Markus Haller, Laurence Kaufmann, Philippe Raynaud, Marc Rüegger and Victoria Tschumi who provided written comments on earlier versions of the dissertation. I have also very much benefited from discussions with Tyler Cowen, Alexandre Dayer, Mark Lance, Maggie Little, James Mattingly and Nicolas Tavaglione.

My work for this dissertation was financially supported by the Ministère Français de la Recherche, the Swiss National Foundation, the Indemnités Chômage, and my husband's income. I am thankful to all of them.

Finally I address my loving gratitude to Alexandre, Ulysse and Agathe, to whom the dissertation is dedicated.

ABSTRACT

The invisible hand is a theory that shows that legislators, collective agreements and moral concerns are not indispensable to the emergence of social outcomes. It instead describes social outcomes as the unintended consequences of many self-interested actions on the part of individuals.

Invisible-hand explanations however often raise criticisms. They are seen as falling short in two ways. First, they unpersuasively present the outcome they explain as an *unintended* consequence of agents' behaviour. Second, they are based on various unrealistic assumptions—that rulers have no impact, that collective agreements do not take place, that agents have no moral concerns —which cast doubt on the explanatory power of their model. The dissertation addresses these two issues.

It begins by elucidating why the class of unintended consequences is likely to fail to fulfil its classificatory role. I then inquire whether invisible-hand explanations require that agents not intend the outcome to be accounted for.

It moves on to explore the explanatory power of invisible-hand explanations. An invisible-hand explanation, according to some of its friends, does not get the facts right because it does not pick out the actual cause of the social outcome. Rather it points out its virtual cause and, hence, explains its resilience.

On an alternative and more promising conception, invisible-hand explanations may well superficially resemble unsubstantiated fiction but they in fact leave aside the accidental features that compose the social realm in favour of its essential elements. It is therefore when one unsuitably expects the invisible hand to fully explain spatio-temporal social facts that one misses its explanatory value.

Expanding on that view, I propose to approach invisible-hand explanations as philosophical explanations, that is, as elucidating the logical structure of social reality. They point out all that is needed for its functioning, neglecting its superfluous elements.

I finally contrast the invisible-hand conception of social reality with Searle's theory of institutional facts. Both understand the institutional reality as one that involves the imposition of functions. But while Searle argues that such imposition needs collective intentionality to be performed, Menger shows how to dispense with the latter. Unlike Searle, however, the relation Menger sees between these functions and the satisfaction of individual interests prevents him from convincingly accounting for the normative dimensions of institutional reality.

INTRODUCTION

By the late Middle Ages, the system of slavery had largely disappeared in Western Europe. What explanation can be offered of this disappearance?

Slavery is an institution whose demise it is quite natural to conceive in terms of abolition, that is, as the project of a legislator. A legislator occupies the position of a central authority and plans the behaviour of a group inasmuch as he holds authority or control over its members. He typically operates by enacting new rules for his subjects. In the present case, the hypothesis is that the Church or the State decreed the end of slavery and punished those who kept on enslaving people instead of hiring them.

The slave owners need not be led by a central authority to do what is required to end the slavery system. They could alternatively have met somewhere and agreed to institute a new behavioural control system. To hire their slaves rather than to coerce them is what they might have collectively agreed to do.

The first explanation is elitist to the extent that it attributes to some ruling mind the ability to invent and implement new institutional rules. The second explanation has a democratic tone inasmuch as it assigns the same idea to ordinary minds. However politically opposed these two explanations may be, they both assert that the end of slavery involved a decision to end it. The extinction of slavery could not have happened, both explanations assume, unless some designing mind had planned it so.

John Millar, a member of the Scottish Enlightenment, offers a different explanation[1]. The end of slavery, he purports to show, need not be a product of an intelligent design. It need not be first represented in someone's mind as a blueprint, and then be put into effect according to some establishment program. Here, rather, is how things evolved. In the beginning, slaves were owned in limited numbers. They were located near their master's house and worked under close supervision. New conquests, however, brought in new slaves, who had to be housed far away. Because they ceased to be closely watched over, the productivity of the slaves decreased, calling for more coercive measures. Owning slaves turned out to be a more costly enterprise than it had been before. Slaves owners thus had to find an alternative way to make slaves work and they came up with the following idea: the slaves would be offered a reward in proportion to the amount of work they performed. Slavery was brought to an end, on Millar's view, when hiring slaves became less costly than coercing them.

1. A surprising explanation

Millar intended to offer a surprising account, one that was at odd with the more intuitive way of explaining the fate of slavery. Its end, he argues, was just the consequence of each slave owners' deliberate decision to pay their slaves. A decision they took with the sole view to secure the productivity of their land.

[1] Millar [1771], 2006, pp. 262-282.

Millar's explanation constitute a departure from the standard explanatory framework which assigns a crucial role to the intervention of a designing mind. On his view, the Church did play a role in the evolution of the slavery system, by reinforcing the institution[2]. No precept of the bible, Millar observes, calls for the abolition of the distinction of ranks or constrains the slave owners' authority. Moreover, slavery persisted in Europe for centuries after Christianity became the established religion, the clergy making it a lucrative source of income for the church. Indeed, ecclesiastical regulations stipulated that every enfranchised slave who obtained his liberty by the influence of the clergy should reward his benefactors. And the laws of the Church also provided that no member of the clergy could liberate a slave without replacing him or her with two other slaves of equal value. Slavery thus disappeared in spite of the support of the religious authorities.

Millar also criticizes those who attribute to the "Great Men" the power to extensively modify the policies of the country that they govern[3]. The decline of slavery, he says, happened "without any express interposition of the legislature"[4]. Legislators are unlikely to abolish slavery because they are "brought up in the knowledge of those natural manners and customs, which, for ages perhaps, have prevailed among his countrymen"[5] and, consequently, endorses "the prejudices derived from ancient usage"[6]. In favour of the *statu quo*, they are unwilling to engage with even shallow reforms.

Millar's account traces back the responsibility for the extinction of slavery to the slave owners, who act in the light of what their personal interest dictates, unaided by any third-party enforcer. More crucially, his account suggests how superfluous such authoritative intervention would have been. Had a central authority endorsed the goal of banning slavery, the slave owners would have been given a further incentive to choose an option, e.g. to pay their slaves, that they quite independently had a reason to find attractive.

A decree is not the only explanatory hypothesis Millar means to dismiss. He also intends to undermine an account that relies on the capacity of agents to engage in collective actions. The end of slavery is not a shared goal, that is, something agents collectively agreed to put an end to. There are no angry slaves purposefully joining forces with a shared view to subverting the system. There are no slave owners collectively agreeing to end the practice of slavery, after having discussed its various advantages and disadvantages. Rather, the end of the slavery system is the result of each slave owners' similar but *uncoordinated* choice.

---

[2] Millar [1771], 2006, pp. 272-273.

[3] The demystification of the legislator is a recurrent theme in the writings of the Scottish Enlightenment. It was already present in Mandeville's *Fable of the bees* ("We often ascribe to the excellency of man's genius and the depth of his penetration, what is in reality owing to the length of time, and the experience of many generations", Mandeville [1723], 1988, p. 142). There is also a passage in the *Theory of Moral Sentiments* where Smith criticizes the "man of system" for believing that the members of the society can be handled as easily as the pieces on a chessboard. "The man of system…is apt to be very wise in his own conceit.... He seems to imagine that he can arrange the different members of a great society with as much ease as the hand arranges the different pieces upon a chessboard. He does not consider that the pieces upon the chess-board have no other principle of motion besides that which the hand impresses upon them; but that, in the great chess-board of human society, every single piece has a principle of motion of its own, altogether different from that which the legislature might choose to impress upon it" (Smith [1759], 1982, p. 42).

[4] Millar [1771], 2006, p. 279.

[5] Millar, [1771], 2006, p. 6.

[6] Millar, [1771], 2006, p. 6.

In the light of Millar's account, collective agreement appears to be as pointless as the intervention of a central authority. The slaves did not need to take the risk of rebelling against the system as the latter progressively ceased to be to their master's advantages. Nor did the slave owners need to engage in any kind of joint action. If hiring the slaves had been in the interests of all the slave owners without being in the interest of any one of them taken individually — if it had been a prisoner's dilemma kind of choice — mutually binding themselves by making the promise to eschew slave labour in favour of paid labour would have been in each slave owner's interest. What Millar shows, however, is that each slave owner has a reason to pay for labour services, regardless of what the other does.

Millar's explanation can be referred to as an "invisible-hand explanation"[7]. We have already stressed two of the features of this type of explanation. Let us consider them in detail and draw attention to a few others.

An invisible-hand explanation is, first, one that dispenses with a designing intelligence. Accounting for the existence of a state of affairs either in terms of the intention of a legislator or that of contractors is what invisible-hand theorists purport to avoid. They describe the possibility that some generally desirable social, outcomes may arise even if agents were not guided through their individual choices by a central authority, on the one hand, or by a binding agreement, on the other hand.

Not all friends of invisible-hand explanations, however, intend to show that *both* a collective agreement and a legislator are superfluous explanatory hypotheses. While some find it important to dismiss the need for rulers, others clearly concentrate on giving a competing and superior account to those which appeal to a collective agreement. Menger clearly abjures both single and collective designers when he presents his own account of the rise of the money system. "Money", he intends to show, "is not the product of an *agreement* on the part of economizing men nor the product of *legislative acts*"[8]. We will here follow him and assume that the designing mind with which the invisible hand dispenses can either be a ruler's mind or a contractual mind.

A state of affairs either is or is not the product of a designing mind. Explanations that refer to the former are "intentional-design explanations"[9]. Invisible-hand explanations and intentional-design explanations are therefore two exhaustive and mutually exclusive options.

A frequent effect of invisible-hand explanation is astonishment. "There is a lovely quality to explanations of this sort", according to Nozick, in that "they show how some overall pattern or design, which one would have thought had to be produced by an individual's or group's successful attempt to realize the pattern instead was produced and maintained by a process that in no way had the overall pattern or design 'in mind'"[10]. In order to be eligible to be an invisible-hand explanation, a social outcome must look like the product of someone intentional design. We must, in other words, be initially misled as to what the good explanation of this outcome is for it to fall inside the scope of an invisible-hand explanation.

---

[7] While Smith invented the metaphor of the invisible hand, Nozick coined the expression "invisible-hand explanation" in the first part of Anarchy, State, and Utopia (1974, see also Nozick, 1994).

[8] Menger [1883], 1985, p. 262 [my emphasis].

[9] Ullmann-Margalit (1978). "Direct explanation" (Wray, 2000) and "intentional explanations" (Elster, 1983 (b)) are alternative designations.

[10] Nozick, 1974, p. 18.

Millar's account is likely to create surprise for one more reason[11]. It is noticeably economical in its commitments. Throughout the process by which slavery comes to en end it is simply assumed that the slave owners and their slaves only aim to improve their own condition. Improving one's wellbeing is, according to Millar, the "powerful motive" that led "our forefathers to deviate from the maxims of other nations, and to abandon a custom so generally retained in other parts of the world?"[12]. This is, in turn, minimally understood as consisting in pursuing the least costly options, among those that they face. The slaves seek to preserve their physical strength. They work to the extent that the costs of not doing so (e.g. being beaten, starved or killed) outweigh the pain of working. "[They] will be idle as often as [they] can with impunity"[13]. The slave owners seek to minimize the costs (in, say, time, money, opportunity and emotional discomfort) of using (coercive or material) incentives. As Millar says, "[they] forced them to labour as much, and gave them as little in return for it, as possible"[14]. It is assumed throughout the explanation that, whether slaves or freemen, agents will not use more resources than they actually need to attain their goal. They are, to put it in more contemporary terms, efficient cost and benefits calculators[15].

Assuming such a cost-benefit calculus on the part of agents certainly has an advantage. It allows Millar to derive a very desirable outcome, namely the end of the slavery system, without assigning any role to beliefs about what one ought morally to do or any other sort of ethical concern. The slave owners are not philanthropic agents who start feeling compassion for the well being of their slaves. They only desire to better *their* own material condition. Were they conscious of the positive effect of material rewards on the well-being of the slaves, the slave owners would represent to themselves such welfare improvement as a mere means toward the enhancement of their own welfare. Since the improvement of the slaves' wellbeing only has instrumental value, it is not any sort of genuine other-regarding consideration. The operation of other-regarding attitudes plays no role in the process that issues in the disappearance of slavery.

In sum, Millar's explanation is surprising because, unlike the explanation that naturally suggests itself, it does not conceive the end of slavery as the result of a process of moral deliberation or growing moral insight. It rejects the romantic idea that slavery was abolished because people became increasingly aware of its cruelty. Slavery was abandoned, he claims, only because slave owners began to find it materially too costly.

Surprise should however not pervade every aspects of invisible-hand accounts. Once our prejudices about the role of legislation, of collective will and of moral sensibilities have been shaken, the ensuing astonishment should give place to a

[11] Not all explanations are surprising. Philosophers often claim that philosophical explanations must not be surprising, in contrast with causal explanations which often are. Invisible-hand explanations are an exception to this rule. They are, I will latter argue in chapter IV, surprising philosophical explanations.

[12] Millar [1771], 2006, p. 262.

[13] Millar [1771], 2006, p. 250.

[14] Millar [1771], 2006, p. 246.

[15] Measuring the costs and benefits of the two behavioral control systems is first the capacity of a few innovators whose personal success is likely to attract many followers. Menger will also resort to a distinction between the pioneers and their followers in his explanation of the rise of money (Menger, 1892a).

reverse sense of "fluency and naturalness", as Ullmann-Margalit notices[16]. This is because although agents involved in invisible-hand accounts act quite unexpectedly, they however do things that neither "extraordinary" nor "freaky" or "on strokes of luck or genius"[17]. Invisible-hand stories would not have any credibility if they were science-fiction kind of story. They therefore assume ordinary people acting prosaically.

2. A parsimonious explanation

Millar depicts a world made of individuals endowed with autonomous, self-interested desires and the ability to choose the options that will best satisfy the latter. Legislators, binding agreements and moral sensibilities do exist. But they are useless cogwheels of the social machine[18]. Dispensing with these oft-invoked explanatory factors is what is distinctive about Millar's explanation. It does not point to additional and hitherto unnoticed explanatory factors. Rather, it explains a lot by a little.

This parsimony is another feature of invisible-hand explanations. Parsimony also is considered as a valuable feature of all explanations. The fewer the number of assumptions an explanation depends on, the better it is. One reason why parsimonious explanations are praised is that they deliver a unified picture of the reality or contribute to providing such a picture. If you can show that not only the end of the slavery system but many other social outcomes can take place without having to assume any capacity to design them in the first place, you will have provided a unified picture of all these apparently disparate social phenomena. You will have shown what is common to all of them, beyond their apparent differences[19].

Also, the world of the invisible hand is parsimonious in a way that makes it easy to implement. The prospect of not having to rely on a central authority to bring about some desirable outcome is good news. As everyone knows, legislators are not entirely reliable good designers. Inventing (good) rules often requires capacities exceeding those of a single agent[20]. Once rules are invented, those who enforce them need to ensure compliance on the part of agents and this involves nudging them. Agents are then no more than the mere instruments or executive organs of the designer's will. They lack the autonomy of free agents because they do not act on their own intentions. They merely execute someone else's plan. Conversely, investigating the possibility of dispensing with a designer amounts to investigating the desirable possibility that agents act fully autonomously. In sum, discovering that no one needs to have designed a rule for it to become operative offers the prospect of not having to count on the mastery of various political skills.

---

[16] Ullmann-Margalit, 1978, pp. 270-271.

[17] Ullmann-Mrgalit, 1978, p. 271.

[18] It is important to stress that although legislators, collective agreements and moral sensibilities are causally inefficient within an invisible-hand explanation, the friends of the invisible hand sensibly recognize them a causal role in the real world and thus do not treat them as epiphenomena. When they say that they are superfluous elements, they mean that their effects would be brought about in their absence and in particular as the result of agents rationally pursuing their self-interested preferences.

[19] Mäki, 1990b.

[20] Hayek, 1973.

What about the prospect of dispensing with collective agreement? The possibility is also attractive for the following reason. Collective agreements have a binding effect. They involve a promise to act in a certain way *no matter what*. Those who make the promise find themselves under the obligation to fulfil it, regardless of what the others do. To be sure, the obligation is not coercive or paternalistic as the obligation imposed by a ruler's authority over those she governs. It nonetheless amounts to a restriction of choice, even if it is a self-imposed one. The invisible hand shows how unnecessary such restriction is.

3. A few other examples

The invisible hand is primarily an economic notion. Its scope, i.e., the sort of phenomena with which it deals, is the economic realm. The invisible hand has been invoked to explain how wealth is redistributed among the less well-off members of society (Smith, 1759), how famine is avoided in period of scarcity (Smith, 1776), how domestic industry gets promoted (Smith, 1776), the rise of the money system (Menger, 1892), or residential segregation (Schelling, 1968). Yet applications of the notion outside its privileged economic domain have also been given. The invisible hand has been given an explanatory role in accounting for many non-economic phenomena such as the emergence of a minimal state (Nozick, 1974), various linguistic changes and even the evolution of language (Marty, 1908, Keller, 1994) and the success of science (Hull, 1988, 1997), among others[21]. It thus illustrates the possibilities of "economic imperialism". In the following, we look at a few of these examples, some of which have been considered as paradigmatic, both within and outside the economic realm.

3.1 The prevention of famine

Although Smith invokes the metaphor of the invisible hand only four times in his writing, the notion is tacitly at play much more often. One such implicit treatment can be found, for example, in a "Digression concerning the Corn Trade and Corn Laws"[22] of the *Wealth of Nations* where Smith discusses the way famine can be avoided during periods of scarcity. The conventional policy was to impose a price ceiling in order to prevent merchants from taking advantage of the shortage by excessively raising the price of the corn. Such intervention will not be necessary, according to Smith, for the reason that merchants will find it in their interest to fix a price that does not overly discourage consumption. They know that if the supply of the season exceeds the consumption of the season, they will have to sell what remains of it for less than they might have had for it several weeks before. Merchants will conversely be interested in fixing a price that adequately discourages consumption so that no shortage occurs before the next harvest. The fate of the population thus depends on a weekly calculation of the price of the corn. It is Smith's contention that the calculation is more likely to prevent the depletion of the

[21] A more complete list would include Hume's description of the emergence private property rules (Hume, [1740], 2000), Ullmann-Margalit's rational reconstruction of the rise of social norms (Ullmann-Margalit, 1977), Gambetta's account of the mafia system (Gambetta, 1993) and Olson and McGuire's explanation of the rise of autocratic regimes (Olson & McGuire, 1996).

[22] I thank Markus Haller for drawing my attention to this example.

grain stock before the next harvest if the people making it have their personal interest in mind rather than that of the population.

3.2 Residential segregation

Schelling is the notorious inventor of a game which models segregation as the result of an invisible-hand process[23]. He asks us to imagine a city as if it were a chessboard, in the form of an 8 by 8 grid square. Its residents come in two dichotomous types: "Blacks and whites, French-speaking and English-speaking, officers and enlisted men, students and faculty, surfers and swimmers, the well dressed and the poorly dressed, or any other dichotomy that is exhaustive and recognizable"[24]. The members of the two groups, say the rich and the poor, are played by dimes and pennies. Initially, the pieces are spread at random over most of the chessboard. It is however essential to leave some squares empty. Each "inhabitant", it is next postulated, prefers to live in a non-segregated environment. Yet each prefers that at least half of his neighbours be of the same social (ethnic, cultural) group as her. Each will be content, in other words, provided half of his neighbours are of the same ethnic group as himself. Inhabitants will continue moving until their preferences are satisfied and the movements are governed by a rule of economy: pieces will stop moving as soon as on their way they find an environment to their liking, that is as soon as they reach the nearest empty square at which their preference are satisfied. Last rule: the game is over when no pieces are left unhappy.

If one moves the pieces according to the stipulated motives, one ends up with a chessboard that displays a segregated pattern. The process is however not straightforward. At some intermediate stages, pieces that were happy at a previous stage will become unhappy and will be motivated to move. Be that as it may, total segregation will inevitably be the end result.

3.3 The minimal state

The next example comes from the first chapters of *Anarchy, State, and Utopia,* where Nozick offers an explanation of the emergence of a minimal state[25]. We first have to imagine people living in the state of nature in the absence of any political authority supervising them. They are equipped with rights, which they defend appropriately against aggressors either by punishing them or by obtaining compensation from them. They may also transfer their rights, such as the right to enforce one's rights, to others. Agents find it in their interest to empower some agencies to protect their rights on their behalf. All agents sign up with one of the specialized protective agencies that emerge during the second stage of the process. It is in the agents' interest not to sign with rogue agencies. That is the reason why all agencies act using reasonable procedures to establish who is guilty, by punishing proportionally, and so forth. During the third stage, the many different agencies are conflated into one single dominant one which offers its service to all agents living in one unified geographical area. Those who in this area have not signed up with any agency, i.e. the "independents", are rightfully prohibited from personally ensuring the protection of their rights. The monopoly enjoyed by the dominant agency in the

[23] Schelling, 1969, 1971, 1978, 1997.
[24] Schelling, 1978, p. 147.
[25] Nozick, 1974. See also Nozick, 1994.

enforcement of rights makes it an ultra minimal state. To become a minimal state, this ultra minimal state must provide compensation to the independents whose rights have been violated when they were prohibited from self-enforcing their rights. The minimal state will compensate the violation of independents' rights by supplying them protection of their rights.

3.4 The success of science

A recent application of the notion is Hull's explanation of the success of science[26]. What Hull wants to explain is why science is so distinctively effective at attaining "its stated goals" which are to give us knowledge of the world. The success of science, he argues, is not the result of the scientists' inclination to discover the truth but rather is the result of the way it is organized as an institution. The institution of the practice of science is more specifically based on a few rules that successfully exploit the scientists' base and selfish motivations so that they end up advancing the interests of science. Hull assumes that scientists do not live up to their romanticized public image, which is to pursue knowledge for its own sake. There is no reason to consider that scientists are different, i.e. more altruistic than other people. As they acknowledge[27], they are mainly interested in gaining credit and recognition from other individual scientists and research groups. The priority dispute is the "most frequent source of discord in science" and reveals that scientists are very far from being motivated by "knowledge for its own sake".

Yet scientists must follow rules, Hull further argues, which will enable them to both reach their selfish goals and to contribute to the success of science. For example, they must use each other's results as well as publish their own results as soon as possible. Those who observe them will distribute the recognition due and obtain the recognition they seek. They will also provide the scientific results that constitute the success of science. Such a coincidence between the scientists' selfish goals and the greater good of science is explained in the following way. "Because scientists must use each others' results and use implies worth, they are forced to give at least some credit where credit is due"[28]. Being constrained to use each other's results is, in other words, likely to produce valuable results. One may however object that using someone's work does not always imply having any concern for its scientific worth. Merely building one's own theory upon someone else's result does not ensure that the resulting theory is valid. Yet Hull believes that it actually will — that "use implies worth" — and his argument is the following. Unworthy results are ultimately filtered out because they will automatically spoil the theory of the scientist who uses them. The filter is one of pragmatic success: the only scientists who succeed are those who have accepted the ideas that really work[29]. Scientists do not have to take pains to test the results of other scientists before using them. Their results are indirectly tested in the sense that any theory that is grounded on non-valid results will itself appear as problematic. Hull claims that, at this point, scientists will want to find where the problem lies. For this purpose, they will trace back the error to the faulty results they have used[30]. In sum, only scientists who use others' valid

[26] Hull, 1988, 1997.
[27] Hull, 1988, p. 309.
[28] Hull, 1997, p. 122.
[29] Hull, 1988, pp. 139-143.
[30] It can here be objected that if scientists in reality hardly care about knowledge for its own sake, why should they be worried about things going wrong with their own research? One possible answer is

results are themselves likely to provide a scientific theory which will itself be valid, a result they only praise as a means to gain the recognition of their peers.

4. Unintended consequences

In the examples considered so far, agents are strikingly unaware of what each of their actions will ultimately bring about. They do not have the outcome to be explained in mind while acting. Of course they are not entirely blind since they have goals and are able to assess whether they achieved them or not. But the goal they individually pursue (successfully in the case of an invisible-hand process) is different from the large-scale social pattern that they produce and the later is the outcome to be explained.

It is an easily recognizable fact that we do not always produce what we intended to bring about. Unintended consequences give a new twist to our actions that could not fail to attract the attention of the most perceptive analysts. The uncontested pioneer in this matter is Mandeville who, in his scandalous *Fable of the Bees*[31], sarcastically reveals how our private vices often and unexpectedly turn into public benefits[32]. Luxurious tastes, he argued, should not be condemned before considering the economic prosperity — by way of increasing the employment rate mainly — they occasion. Hume also offered insightful observations about thwarted plans. Both his *Essays* and his *History of England* are riddled with statements of perverse effects, especially of the kind rulers awkwardly cause when trying to preserve their power. Yet pride of place in this tradition must be accorded to Smith. He famously notes in the *Wealth of Nations,* that the entrepreneur "promotes an end which was no part of his intention"[33] by safely investing his money in domestic industry, whose commercial rules he is familiar with, rather than abroad. Echoing Mandeville, wealth is described in his *Theory of Moral Sentiments* as what the rich Landlord ultimately finds himself sharing with those he hires to make the baubles and trinkets he is so idiotically fond of. Millar in turn applied the same idea in order to cast a new light on the end of the slavery system, which he saw as the side effect of a calculated choice on the part of the slave owners when they turned to material incentives instead of coercion as a behavioural control system[34]. The most eloquent and oft-quoted statement of the idea, however, comes from Ferguson[35] who noted how "nations *stumble upon establishments, which are indeed the result of human action, but not the execution* of *any human design*"[36]. Burke deserves credit for drawing the political conclusion that follows from viewing the social order as an unintended consequence of various prejudiced choices[37]. If things stand as they do without any deliberate

---

that they will be afraid of having their faulty results being exposed by other scientists. Exposing someone's faulty results must however be done for the sake of redirecting the credit to oneself, rather than with for the purpose of contributing to scientific discoveries.

[31] Mandeville (1988).

[32] Hayek, 1966. The role that Mandeville attributes to a "skilful politician" in his ability to "extract good from the very worst", however, raises doubt as to whether the outcome is fully unintended.

[33] Smith [1776].

[34] Millar 1771.

[35] For a complete analysis of the place of unintended consequences in Ferguson's thought, see Hill (1998).

[36] Ferguson, 1782, p. 187.

[37] Tieffenbach, 2002.

effort on our part, he argued, proceeding to a reconstruction of that social order from scratch is likely to produce more damage than goods[38]. Hegel's idea of the "cunning of reason" is the idea that men of the greatest kind are the blind tools by which Reason (or whatever supra-individual force) unfolds its plan[39]. Insightful remarks about the conditions for political success and defeat abound in Tocqueville's *Souvenirs*. As Holmes shows[40], the paradoxes of being "saved by danger and destroyed by success" are the book's main hypotheses. An idea Tocqueville poetically expresses as follows: "Sur la destinée de ce monde qui marche par l'effet, mais souvent au rebours des désirs de tous ceux qui la produisent, semblable au cerf-volant qui chemine par l'action opposée du vent et de la corde"[41].

Until then, unintended consequences were the sort of effects that those occupying the highest ranks in society either incompetently and/or happily brought about. Menger can be credited with showing that common people were no less involved in manufacturing these by-products. Many institutions, he notes, "are not the result of socially teleological causes, but the unintended result of innumerable efforts of economic subjects pursuing *individual* interests"[42]. The novelty of Menger's analysis was also to explore in unmatched detail the path by which social facts could unintentionally emerge. His fictive genealogy of the money system[43] shows how it could prevail within a group just as a consequence of its members seeing the benefits they individually could draw from acquiring highly marketable goods for the purpose of exchanging them in the market place.

Menger set the tone: to advance an invisible-hand explanation is to depict many common minds who do not have the ultimate result of their action in mind while acting. Another way to describe the same constraint is to say that the outcome must appear as the unintended consequence of many actions that are directed toward other ends. Unintendedness is thus a feature displayed by all outcomes that have been invisible-handedly explained. It is a constraint visible in the various applications of the notion.

Unsurprisingly, this constraint also appears in most clarifications of the notion. As Mäki claims, the outcome to be explained must fall outside agents' "sphere of intention"[44]. An invisible-hand process is regularly presented as involving many agents who are too embedded in their individual position to see the aggregative effect of their actions. Tuomela summarizes an uncontested and widespread view when he defines an invisible-hand explanation as one that "replaces an easily forthcoming and initially plausible explanation according to which the *explanandum* is

[38] In his review of this reactionary rhetoric, Hirschman famously calls "the perversity thesis" the argument whereby any action to improve the political, social, or economic order will counterproductively result in the opposite of what was intended (Hirschman, 1991).

[39] Tieffenbach, 2003. The authoritarian, elitist, and theological dimensions of Hegel's cunning of reason, however, hardly bears comparison to the Scottish notion of unintended consequences in social reality. The following passage taken from Hegel's correspondence speaks for itself: "Je m'en tiens à ceci: l'esprit mondial de notre temps a donné l'ordre d'avancer. On obéit à un ordre. Cet être avance comme une phalange cuirassée et solidement soudée, irrésistiblement, avec un mouvement aussi imperceptible que celui du soleil, envers et contre tout, en avant… La plupart ignore de quoi il [en] retourne, et se contente de recevoir sur la tête des coups qui semblent venir d'une main invisible" (Hegel, 1975, p. 28).

[40] Holmes (2009).

[41] Tocqueville, 1978, p. 66.

[42] Menger, 1985, pp. 147, 158.

[43] A description of such genealogy is given in introduction to chapter IV.

[44] Mäki, 1990a.

the product of intentional design with a rival account according to which it is brought about via a process involving the separate actions of many individuals who are supposed to be minding (only) their own business unaware of and hence not intending to bring about the ultimate overall outcome[45]. An invisible-hand explanation involves agents whose aims are not coordinated nor identical with the actual outcome, which is a by-product of those aims. The process works without the agents having any knowledge of it.

Take, for example, Millar's explanation of the end of slavery. Not only is the abolition of the slavery system no one's goal, it is an aggregative result that no one, not even the slave owners, had in mind while acting. The rooting out of slavery is in other words the unintended consequence of several actions that are directed toward other ends. Similarly, in Nozick's explanation, no one needs to intend to create a state for a state to arise. One protective agency becomes dominant because individuals choose to sign up with the agency that already provides the best protection, not because they intend it to be an ultra minimal state. In its turn, this ultra minimal state becomes a minimal state because those who manage the dominant agency meet their obligation to compensate independents, not because they want their dominant agency to qualify as a minimal state.

On the invisible-hand conception of social reality, every significant thing that happens in the social world happens behind people's backs. Though the invisible hand assigns no role to Great Men, it conversely concedes no more deliberate power to the great unwashed. The latter are too blind to bear any responsibility in the way the social world works.

Uncovering the unintended consequences of an agent's action is a rewarding activity. Perhaps the pleasure comes from offering an unusual view of how the world functions, one that defies the all too common tendencies to see it as either a natural or an altogether designed order[46]. More plausibly, the pleasure comes from offering a description of social reality that the very agents who make it possible cannot themselves provide.

However enjoyable describing a social outcome as a by-product might be, defining the invisible hand as being about unintended consequences raises a few problems[47]. Only outlined here, these problems will be discussed at length through out the various chapters of the dissertation.

Firstly, it entails a certain degree of blindness on the part of agents that is, at times, close to idiocy[48]. Often the relation between what agents individually do and what they, as a group, bring about, is so obvious that their lack of awareness appears

[45] Tuomela, 1984, p. 451.
[46] Hayek, 1978.
[47] In a (much debated, cf. Aydinonat, 2006) chapter of her *Economic Sentiment* (2001, chapter V), Rothschild provocatively argues that the invisible hand is "the sort of idea that Smith would not have taken entirely seriously" (Rothschild, 2001, p. 118). One reason is that the condition of blindness that is at its core is at odd with Adam Smith's idea of human nature. The invisible hand is "un-smithian", she also argues, because of its elitist aspect. "It presupposes the existence of a theorist … who sees more than any ordinary individual can… The disembodied hand is invisible to its millions of petty subjects, but it is visible to 'us': to theorists." (Rothschild, 2001, p. 123). It may however be replied that someone well acquainted with the division of labour like Smith would not be troubled by the possibility that those who attempt to understand the social world could acquire a more accurate understanding of the latter than those who spend their time differently.
[48] Zaibert, 2004.

as mere stupidity. One cannot restrict the application of invisible-hand type explanations to only societies of fools.

Secondly, defining an invisible-hand consequence as an unintended consequence inconveniently leaves most explanations in social sciences unclassifiable. This is because it is often impossible to determine how far from agents' sight the consequence must be situated to count as an unintended rather than as an intended one[49].

Thirdly and crucially, defining the scope of the invisible hand in terms of unintended consequences rules out some intuitively relevant cases. It does not, in particular, accommodate the cases where agents knowingly manage to coordinate their actions by intentionally adopting the same social rule, unaided in this task by speech or by any external coordinator[50].

5. A potential explanation

Let us now turn to another noteworthy recurrent feature of invisible-hand accounts of the social realm. No evidence, as any historian will have noticed, is given in support of their truth. Consider in this regard Millar's explanation of the end of slavery. He strikingly does not refer to any archive that would testify that the slave owners really reasoned in the lucid and self-interested manner that he describes. He does not mention any of the supporting documents simply because he does not have access to the information that would have told him whether his explanation is true, partly true only or even false. In other words, Millar seems to be exploring a mere possibility, using mainly his imagination, rather than to be describing how things really happened.

Millar is not in this respect guilty of any kind of methodological sin. In fact, his explanation is a good example of what his peer Dugald Stewart (1753-1828) refers to as "Conjectural History"[51]. As Stewart explains, conjectural history stems from the lack of evidence concerning the event one wishes to explain. "When it is impossible to reconstruct the process by which something has happened", he claims, "it remains useful to show how it could have happened"[52]. Although a would-be cause is likely to differ from the actual cause of a particular phenomenon, it nonetheless is worth being stated. Speculating this way, Stewart argues, will at least cast more light on the phenomena to be explained than invoking a miracle.

Conjectural history is still a widespread activity. In fact, most paradigm invisible-hand accounts fall into this category. An invisible-hand explanation, it is often pointed out, is a conjectural account of how a certain social pattern might have come about. It characteristically remains silent as to how this pattern actually arose. Invisible-hand explanations are variously described as "potential explanations"[53], as

[49] Vernon, 1978, Cowen, 1997.
[50] These cases are discussed by Lewis (1969) and Schelling (1960) to be presented in chapter II.
[51] Stewart, 1858, p. 34. Stewart's idea of conjectural history anticipates the genealogies discussed by Nietzsche [1887] (1994), Craig (1990) and Williams (2002).
[52] Stewart, 1858, p. 34.
[53] Nozick, 1974, Ullmann-Margalit, 1978.

"conjectural histories"[54] or as "how-possibly stories". It is thus a common view that an invisible-hand explanation is an explanation of how things *could* happen.

A potential explanation is, on Nozick's definition, "what would be the correct explanation if everything in it were true and operated"[55]. Consider, as an illustration, his explanation of the rise of a minimal state. Nozick expressly denies any truth to his account. It is, as he says, a "defective"[56] account. Certainly, history displays cases of stateless groups within which the establishing of a central authority took place. But real central authorities are likely to have been produced by some other processes — e.g. a process during which the weakest were subjugated by the strongest — than the one Nozick depicts.

What makes Nozick's account look like a just-so story is the nature of the various assumptions on which it is based. Nozick assumes that agents initially lived in a state of nature, that they act rationally and that they "satisfy moral constraints and generally act as they ought"[57]. These are rightly considered as unrealistic assumptions.

The same can be said of all other invisible-hand explanations. They all are based on assumptions that do not describe the reality accurately. As some have noticed[58], the most telling example of all is, in this respect, Schelling's checkerboard city. Schelling never intended to offer an historical description of the process by which segregation arises in real cities. His checkerboard model merely points at a *possibility*. It shows how, under some conditions, mild segregationist preference causes total segregation. These postulated conditions include the following: (i) pawns are initially randomly distributed, (ii) pawns have mildly segregated preferences about the composition of their neighbour, (iii) pawns move to the nearest available place that satisfies their preference, (iv) pawns move continuously until their preference is satisfied. These conditions are however not accurate descriptions of reality. Residents are not initially randomly located, nor can they always move to other places. The assumption that they have the same relatively tolerant preference about their neighbours is also unfounded. The checkerboard city is thus a conjectural account of segregation.

Of course an invisible-hand explanation may turn out to be true. Even then, it would however be only accidentally true. It would be true as the result of describing the world in a way that *incidentally* happens to be truthful. Its truth would not be the result of tracking reality and the reason is that truth does not seem to be what friends of the invisible hand are preoccupied with. An invisible-hand explanation is not valued in virtue of its ability to describe the reality accurately. Invisible-hand explanations are valuable, it is argued, "even if they do not get the facts rights"[59]. One way to spell out this view is to say that an invisible-hand explanation is valuable in virtue of its ability to meet various internal constraints. We previously mentioned two of them, namely parsimony and the capacity to elicit surprise. "Sophistication"[60] and "a certain lovely quality"[61] have also been added. However, these happen to also

[54] Keller, 1994, Aydinonat, 2008.
[55] Nozick, 1974, p. 7.
[56] Nozick, 1974, p. 7.
[57] Nozick, 1974, p. 5.
[58] cf. Sugden, 2002, Rosenberg, 2007, Aydinonat, 2008.
[59] Aydinonat, 2008, p. 6.
[60] Ullmann-Margalit, 1978, p. 274.
[61] Nozick, 1974, pp. 18-19.

be qualities of short stories and the latter are usually not interchangeable with explanations. As Elster says, "explanations must be distinguished from storytelling. A genuine explanation accounts for what happened, as it happened. To tell a story is to account for what happened as it might have happened (and perhaps did happen)… why would anyone want to come up with a purely conjectural account of an event? Is there any place in science for speculation of this sort? The answer is yes — but their place must not be confused with that of explanations"[62]. Invisible-hand explanations are precisely the sort of storytelling that Elster criticizes. Intuition tells us that learning how a social pattern could possibly come about is not the sort of information that we look for when we ask for an explanation of that given social pattern.

Notwithstanding their conjectural character, the explanatory value of invisible-hand accounts is regularly defended. Nozick, for example, finds special interest in conjectural histories. He says that his explanation "will serve our explanatory purposes" "even if no actual state ever arose that way"[63]. On Nozick's view, "we learn much by seeing how the state could have arisen, even if it didn't rise that way"[64]. An invisible-hand explanation, others argue, "expands our horizon" and is "interesting or 'illuminating' in its own right"[65]. What kind of explanatory insights do we gain by being told a story about how things could have happened if we know that things did not actually happen this way? What do we precisely learn by means of a how-possibly explanation? If it is true that a would-be invisible-hand explanation "carries important explanatory illumination"[66], what exactly it highlights and how it performs such a function still needs to be examined.

6. An alternative account

An invisible hand explanation of a social outcome is a conjectural account describing it as what agents would unintentionally bring about, supposing they were only interested in their material welfare. As we have seen, any attempts to explain social reality in this manner raises problems of various kinds, which ultimately put into question the explanatory power of such an account. What is it that we get knowledge of by means of an invisible-hand explanation? What if anything is explained by describing a social outcome as a would-be unintended consequence? The dissertation offers the following general answer to this question.

On the proposed account, invisible-hand explanations are not causal explanations, although they certainly very much look like the latter. Rather they are philosophical explanations. Invisible-hand explanations illustrate a type of inquiry that purports to highlight the essential components of social reality. They are conceptual investigations that aim to discover the essence of institutions. They elucidate what we mean when we speak of money, the state, language, autocratic regimes, redistribution, or of the mafia system, they elucidate what sort of things these are. They present the sufficient conditions under which a social outcome (whether it be a social system, a social structure or a social pattern) is money, is a

[62] Elster, 1989, p. 7.
[63] Nozick, 1974, p. 7.
[64] Nozick, 1974, p. 9.
[65] Ullmann-Margalit, 1978, p. 275.
[66] Nozick, 1974, p. 8.

state, is a segregated city, etc. An invisible-hand explanation, i.e. an imaginary reconstruction of a process resulting in the emergence of a given social institution, will highlight the sort of dependency that links the social realm to the psychological and to the physical realms.

*

The structure of the five chapters of the dissertation is as follows.

Chapter I discusses the standard characterization of invisible-hand explanations in terms of unintended consequences. It first offers a detailed version of this characterization, using the so-called description theory of action. Describing an outcome as the unintended consequences of agents' actions is to give an accurate description of these actions under which their authors did *not* act. The chapter then deals with various puzzles that arise from such a characterization. It first inquires about the sorts of description that agents who are led by the invisible hand are prohibited from offering, or permitted to offer, of their action. It clarifies in particular the common idea that these agents are supposed to lack the theoretical knowledge that would allow them to offer such description. I then examine whether there are any social outcomes that (i) *can only be* the consequences of many actions directed toward other ends?" or that (ii) *cannot* be the consequences of actions directed toward other ends. Answers to these questions help to delineate the scope of invisible-hand explanations. They indicate whether there are some social patterns that necessarily fall within their scope and whether there are, on the contrary, social patterns that necessarily escape them. I conclude that defining the invisible hand in terms of unintended consequences of actions makes obscure rather than illuminates the distinction between invisible-hand explanations and intentional-design explanations.

In **chapter II**, I investigate the extent to which the outcome can enter the agents' "sphere of intention"[67] without ruling out the invisible hand. I address this question by submitting the standard account to various revisions until a case is found that clearly lies outside the bounds of the invisible-hand type model. My conclusion is that an invisible-hand explanation may very well welcome agents who have the intention to produce the outcome to be explained. It may very well accommodate intentionality if it remains the *individual* form of intentionality that gives shape to individual actions. What it excludes is, in other words, the collective form of intentionality that is specifically at work in social pact stories where agents *jointly* reflect on the social rules they should all conform to. On the proposed view, the scope of invisible-hand explanations must be re-defined in a way that allows agents to intend, albeit *individually*, to do what is needed for the outcome that is to be explained to obtain.

The following two chapters examine how an invisible-hand explanation can be explanatorily valuable in spite of its imaginary character. Two solutions in particular are considered. Invisible-hand explanations are first said to explain the robustness of social patterns. An invisible-hand explanation, it is argued, does not account for the way a social pattern emerge, or for the way it maintains itself. Rather it explains why

[67] Mäki, 1990a.

it is likely to persist[68]. This is because an invisible-hand explanation should not be treated as referring to the actual cause of the outcome. It should be approached as indicating its "virtual cause" — to what would be its cause, were its actual cause not operative. This solution will be discussed in **chapter III**.

The second solution, to be pondered in **chapter IV**, was advanced by Carl Menger more than a century ago[69]. In a nutshell, Menger argues that invisible-hand explanations are not potential explanations because they describe the reality, but because they describe it in a highly partial way, representing reality incompletely, leaving aside many of the features that compose the social realm, rather than inaccurately. Menger then argues that the elements of reality that are deliberately neglected are the elements that pertain to what he calls an "empirical" approach of to reality. By contrast, the elements that are retained are those that are relevant to what he calls an "exact" understanding of the same reality. He then argues that it is only when one inappropriately expects an exact account to provide the sort of explanatory information that only the empirical approach can give that one misses the explanatory value of the former. Menger's solution is appealing as it claims that, contrary to what we said earlier, friends of the invisible hand do take truth as their central concern. I will however show that it does not persuasively defend the rational choice assumptions on which invisible-hand explanations are based.

Chapter V presents Searle's conception of institutions in terms of status function in order to highlight its similarities and differences with the invisible-hand conception of institutions. Taking Menger's view as illustrative of the invisible-hand theory of social reality, I first show that the assignment of (what Searle refers to as) status functions plays a crucial role in both theories. I then argue that whereas Searle holds that status functions can only be collectively assigned, Menger denies the need for collective intentionality in the ascription of function. A second divergence between the two conceptions concerns the role they each attribute to the satisfaction of desire in the existence of institutional facts. Whereas Menger takes it to be an essential component, claiming that all institutions should be approached as satisfying agents' desires, Searle shows what is wrong with this conception, that is, its failure to account for the normative dimensions of institutional facts. Indeed, the rights, duties, obligations, commitments, authorizations, or requirements that one is subjected to when one orders a beer, uses money, goes to university, drives on the road, or attends a cocktail party are unrelated to what one may or may not desire to do. Rather, we have a desire-independent reason to recognize the various status functions on which all institutions partly depend.

[68] Ullmann-margalit, 1978, Pettit, 2008.

[69] Recent reconstructions and defenses of Menger's argument can be found in Smith (1986, 1995, 2010) Mäki (1990a, b, c, 1997), Haller (2002, 2004), Nadeau (2003, 2005) and Aydinonat (2008).

I - EXPLAINING THE UNINTENDED CONSEQUENCES OF ACTIONS: A FEW CONUNDRUMS

Introduction

According to all proponents of methodological individualism, offering an explanation of a social outcome requires that we penetrate agents' intentions. Doing so will cast light on the success of these intentions. Either the social outcome matches or fails to match these intentions, respectively giving occasion to an intentional-design explanation or to an invisible-hand explanation.

The distinction between intended and unintended consequences has therefore classificatory power[70]. We can, on its basis, distinguish between two jointly exhaustive and mutually exclusive types of explanations.

The distinction seems at first sight to be clear enough to serve this purpose well. There does not seem to be any problem in assessing whether a given outcome is described as being intentionally brought about or not.

Yet friends of the invisible hand often are criticized for not providing the invisible-hand kind of explanation that they intended to offer and their failure, it is argued, is to unpersuasively present the outcome to be explained as an *unintended* consequence of agents' behaviour. In other words, many so-called paradigm invisible-hand explanations do not, on close scrutiny, deserve their labels. Taking Hayek's interpretation of the Scottish Enlightenment as her target, Petsoulas[71] argues along this line. She claims that Mandeville, Hume, and Smith's various accounts of institutions do not qualify as invisible-hand explanations because they assume too many reflexive and purposeful aptitudes. On her view, the central role that the Scottish Enlightenment assigns to conscious reflections and imitation in the development of institutional rules is incompatible with the supposed unintended

[70] Showing how things do not turn as we thought they would is considered as a path-breaking step in the history of social thought. On Hayek's view, describing social outcomes as the unintended consequences of human's actions no less contributed to the rise of "the modern mind" (Hayek, 1978, p. 190). Popper regards the discovery of unintended consequences as the "main task of theoretical social sciences (Popper, 1962, p. 342). Social science could not emerge as an independent discipline, Hayek also argues, until the notion of unintended consequences had been clearly envisioned. For this we had to stop viewing the world as either a "natural order" or as "an artificial order". Unlike natural orders, artificial orders are all action-dependent, as someone needs to do something for them to be obtained. But that action-dependency may, in turn, be diversely conceived. It might either be dependent on intention or not. Recognizing the latter possibility will make the third category of "spontaneous orders" apparent:

Natural order	Artificial order	Spontaneous order
Independent from human action	Depends on human action	Dependent on human action
Independent from human intention	Depends on human intention	Independent from human intention

[71] Petsoulas, 2001.

character of these rules. Their explanations should therefore be reappraised as intentional-design explanations[72].

Menger's account of the emergence of the institution of money is another example of an allegedly misnamed invisible-hand explanation. Thus Steiner argues that the emergence of the money system would not be conceptually possible if people did not collectively agree to establish it. "The establishment … of a common medium of exchange", he claims, "cannot be construed as the effect of the operation of invisible-hand processes. Rather [it] must be attributed to the deliberate and concerted efforts of individuals to bring [it] about. [It] must be attributed to social contracts"[73]. The argument is the following. Steiner first argues that money is not a more efficient system than the barter system. He first (painstakingly) demonstrates that the use of a common medium of exchange is actually not more efficient than barter. The number of transactions required to clear the market, according to Steiner, turns out not to be lower under the money system than under the barter system. Moreover, money is not more portable than some goods if barter transactions include the exchange of vouchers. A reason for accepting money in exchange for one's goods is the belief that monies are goods that will produce the goods I want when I want them. However, the availability of these goods does not depend on me. I believe these goods will later be available if I believe that the others who supply these goods will also prefer to exchange them for money rather than for other goods at that time. Hence, "the acceptability of money falls within that perplexing but fascinating group of phenomena which is affected by self-justifying beliefs. If members of a community think that money will be generally acceptable, then it will be"[74]. Otherwise, it will not. Therefore, the challenge is to explain how agents acquire the belief that money will be generally acceptable. Steiner regards it as a prisoner's dilemma type of situation. Agents will not acquire this belief, he argues, unless they commit themselves to go against their self-interest. Refusal to accept the other's money is the dominant strategy. If I am unable to rely upon the other's acceptance of my money, I will rationally prefer to retain my own goods rather than relinquish them in exchange for the other's money. To rationally prefer the latter, I need to be under the pressure of a pre-commitment. Pre-commitment is what I am required to do as a party to a social pact. Therefore, a common medium of exchange is a public good, which "presupposes a general contractual understanding among individuals to accept money in exchange for goods"[75].

According to Steiner, Nozick's explanation of the minimal state lends itself to the same objection. A monopolistic protective agency, he argues, could not emerge if agents did not entertain, believe, want and desire it to emerge. To grasp this point, let us consider a situation where many protective agencies are competing in order to attain a monopoly. Each supplier must know that his agency will not survive if it does not secure the largest possible number of clients. Since the "largest possible number of clients" means "all clients", suppliers may be said to believe, want and desire that a single agency monopoly be brought about. As for the clients, "they each

[72] The objection should not be confused with an altogether different way of rejecting the need of an invisible-hand explanation, that is, in terms of its empirical inadequacy. Wray's objection to Hull's account of the success of science is of that type (Wray, 2000) when he blames Hull for not recognizing the scientists' intentions to seek knowledge of the world. This way of criticizing the relevance of an invisible-hand explanation is discussed in chapter IV.

[73] Steiner 1978, p. 295.

[74] Steiner, 1978, p. 312.

[75] Steiner, 1978. p. 312.

know that it is in their interest to contract with that supplier who contracts with the largest possible number of persons. It is in no one's interest that others patronize agencies other than the one the client patronizes. If fewer persons contract with one supplier rather than with another, not only are the clients of the former worse off than those of the latter, but both sets of clients are also worse off than they otherwise might be. These clients too will believe, want and desire that a single agency monopoly be brought about. … Clearly it is in the perceived interest of each person to agree with every other person as to which agency they should all patronize"[76].

Gaus also disputes Nozick's explanation on similar grounds[77]. On his view, it is not obvious that the rise of a minimal state is not among the goal of the dominant protective agency. "The actions and reasoning of the dominant protective agency in prohibiting unauthorized enforcement by independents", he notices "is too close to aiming at [the minimal state] to constitute a satisfying invisible hand explanation of [the mininal state]: the agency is seeking to gain a monopoly on the authorization of coercion, and its complex compensation reasoning is anything but prosaic. It is not very surprising that the outcome of the dominant agency's reasoning is a claim to minimal, Lockean, statehood"[78]. Because the attempt to gain a monopoly on the authorization of coercion is not truly different from the intention to aim at a minimal state, the transition from an ultra-miminal state to a minimal state seems more deliberately brought about than what it is supposed to. Nozick's explanation of the minimal state is therefore too closed to a social pact explanation of the same social phenomenon to be a compelling illustration of an invisible-hand explanation.

The goal of this chapter is not to defend the validity of these objections nor, for that matter, to respond to them. Rather it is to show why the standard definition of the invisible hand is likely to give rise to them. At the core of Petsoulas and Steiner's argument is the idea that, although invisible-hand explanations purport to describe the outcomes to be explained as unintended consequences, they conspicuously fail in this task. The reason is that they involve agents who are not sufficiently minding their own business and thus appear to some extent as seeking to realize the outcome, turning it into an intended consequence of their actions. The gap separating the agents' individual intentions and what they bring about in the aggregate, Petsoulas and Steiner argue, is not wide enough to allow the working of an invisible hand.

Classifying an explanation on the basis of its ability to describe its object phenomenon as an (un)intended consequence is likely to trigger disagreement. My way to approach the problem is to examine whether the intended *vs.* unintended consequences dichotomy can be disambiguated, *viz.* to what extent its sense can be sharpened so that it proves to be an adequate classificatory tool.

Here is how the different sections of this chapter address this question. Section 1 clarifies the distinction between unintended and intended consequences, by using a conventional theory of intentional action. On such theory, describing an outcome as the unintended consequences of agents' action is to find an accurate description of the behaviour under which their performers did not act. Section II inquires about the nature of the description that agents who are led by the invisible hand are prohibited from offering (or permitted to offer) of their action. It examines in particular the idea

[76] Steiner, 1978, p. 314.
[77] Gaus (forthcoming).
[78] Gaus (forthcoming).

that agents are supposed to lack the theoretical knowledge that would allow them to be aware of what they are doing. The shortcomings of this view will be exposed in section III. As an alternative approach, section IV inquires whether there are any social outcomes that *can only be construed as* the consequences of many actions directed toward other ends. Section V by contrast explores whether there are any social outcomes that *cannot* be construed as the consequences of actions directed toward other ends.

1. Acting under a description

1.1 The sphere of intendedness

On all account, the notion of an unintended consequence is an essential ingredient of the invisible hand. The notion presupposes a distinction between consequences that agents bring about intentionally and consequences that they do not intend to produce but produce all the same.

The notion of unintended consequences is however an intricate one. It is not easy to tell where to draw the dividing line between intended and unintended consequences of agents' actions. Mäki offers a framework for adjudicating cases by means of a distinction between what lies inside and outside the "sphere of intendedness" of someone's action[79]. The latter refers to "the states of affairs that an acting agent intended to bring about by his or her act"[80]. The idea is simply that "every intended result for an action belongs to the sphere of intendedness of this action"[81].

Mäki further notes that "the same type of behaviour may be accompanied by various spheres of intendedness"[82]. To illustrate such possibility, he gives the example of someone, called A, turning the handle of a window. This event, turning the handle of a window, could in some instances be the only event that belongs to A's sphere of intendedness. But had A intended to open the window, the sphere of intendedness would accordingly be larger. The volume of the sphere could even include other events such as finding out what made the strange noise from the street, or getting some fresh air into the room. Suppose that these events exhaustively fill in the sphere of intendedness of A's behaviour. This does not prevent A's action of turning the handle from having many of these other ulterior consequences. The heating cost of the building, the drop of temperature, D sneezing as he happens to have a flu, A being contaminated by D's flue, A being unable to recover soon enough to give her presidential address at the Annual Meeting of the Society for interpretive Economics. None of these consequences, however, are within the sphere of intendedness of A's act. Nor can they, therefore, be accounted for by referring to A's intentions.

1.2 Actions are intentional under descriptions

[79] Mäki, 1990a, pp. 161-163.
[80] Mäki, 1990a, p. 302.
[81] Mäki, 1990a, p. 161.
[82] Mäki, 1990a, pp. 161-162.

Yet A's action can be described as one that raises the heating cost of the building, decreases the temperature, makes D sneezing and so on. That A does not have the intention to bring about these various consequences does not alter the fact that producing these consequences is something that she nonetheless does[83].

Commonly stated, bringing about these consequences is something that A did unintentionally. We can spell this out in the following way. When A opened the window, she engaged in an action *under some descriptions* she knew or believed apply to her action[84]. She acted under the description "opening the window", as well as under the description "finding out what made the strange noise she had heard from the street". These descriptions specify the particular perspective from which she acted. Being in this particular position, and not, for example in D's position, A had limited knowledge of what it was that she did. The position from which she acted allowed her only a partial view of what she did.

From D's perspective, however, A did something else (in addition to opening the window and inquiring about the noise she heard). She made him sneeze. What A did from D's perspective is something A really did, although unintentionally. The usual example of such a scenario is Oedipus. His action was intentional under the description "marrying the queen", but not under the description "marrying his mother". Oedipus acted from his own perspective, according to which 'being the queen' was true of the person he married. Either Oedipus mistakenly believed that "marrying his mother" was not true of himself, or he had no belief whatsoever on the question.

"Married his mother" and "married the queen" are two accurate descriptions of Oedipus' action. But only the second description can finish the sentence: "Oedipus intentionally …". This is because having an intention is like having a desire or a belief. It is a state of mind that is referentially opaque. Truth is not always preserved if we replace one description with another description of what is intended. In other words, descriptions of the contents of intentions are "intensional".

We do not know all the descriptions that apply to our actions. Yet there is nothing wrong with referring to what someone does using a description of which he is not aware. As we said, we will be in this case attributing to her an action that she does, albeit unintentionally.

In sum, what counts as doing something intentionally and unintentionally depends on how the agent conceives of what he does. Or, as Moya says, "in order to attribute intentionality to an action it is essential to take into account the description of the action under which intentionality is attributed."[85]

Let us consider Nozick's explanation of the minimal state in light of these remarks. Nozick makes it clear that the situation in which agents find themselves when they have all subscribed to the monopolistic protection agency can rightly be described as one in which a minimal state exists. But establishing a minimal state is something that none intended to do. Nozick stresses the point in the following way. "We should note that ... the explanation... does not specify people's objective as that

[83] Note that this only holds if the unintended consequence is not too remote. Suppose, for example, that opening the window has the consequence of severely aggravating D's condition. Endangering D's life would be a consequence of A's opening. Yet we would resist describing it as something that A does and the reason seems to be that it is a too distant effect of A's opening the window.
[84] Anscombes, 1957, §§ 26-29.
[85] Moya, 1990, p. 52.

of establishing a state. Instead, persons view themselves as providing particular other persons with compensation for particular prohibitions they have imposed upon them"[86]. Not only is the establishment of a minimal state no one's aim, it is something that happens without agents even knowing it. Let however agents describe what they do as establishing a minimal state, however, and Nozick's explanation will have to be re-labelled as an intentional-design explanation.

All invisible-hand explanations can be expected to lend themselves to the same sort of elucidation. The idea is to look at these examples in light of the following two questions. Under what descriptions of their actions do agents act? What descriptions are, on the contrary, unknown from their perspective?

Take, for example, Millar's explanation of the slavery system. None of the slave owners intended to "to play a role in the extinction of the slavery system". Rather they would describe their action as one that "embraces the salutary policy of *bribing* [the slaves], instead of using compulsion, in order to render them active in their employment"[87].

Take, as another example, Menger's account of the emergence of the money system. Pick anyone involved in the process described, and call him A. It seems right to say that A intends "to acquire a highly traded good" and that A intends "to acquire a good he does not necessarily presently need". These are, by hypothesis, descriptions that, from A's perspective, apply to his choice.

Would Menger further agree to say that A intends "to exchange his less traded goods for a more marketable one"? Probably not. Menger rightly expects most agents to ignore what a highly marketable good is since they do not master the concept of "marketability". Surely, they know what a highly traded good is, but they do not know that a highly traded good is a highly marketable good. By the way, there are other descriptions that probably escape A's particular perspective but that are true descriptions of A's action. For example, A does not intend to "increase the popularity of the highly traded goods that he attempts to acquire". Nor does A intend to "contribute to the establishment of the money system". Even more remote from his mind is the idea of "acting in accordance with the public interest".

Consider next Smith's explanation of the distribution of wealth. Call A the rich Landlord whose own wealth ends up benefiting many of his employees. Among the descriptions Smith offers of A's action, some would be recognized by A as applying to his choice while others would not be. A can certainly be said to intend "the gratification of his own desires"[88]. Another description under which A acts is "to mean only [his] own conveniency". But there are many descriptions of A's action that escaped A's mind. Examples are "to divide with the poor the produce of all [his] improvements", "to advance the interest of the society" and "to make nearly the same distribution of the necessaries of life, which would have been made, had the earth been divided into equal proportions among its inhabitants".

Consider finally Schelling's checkerboard city. Although A intends "to move into a neighbourhood in which at least half of her neighbours is of her colour", he does not intend "to affect the environment of those he leaves and those he lives next to"[89]. A intends to "keep moving until his preference for mild segregation is

[86] Nozick, 1974, p. 119.
[87] Millar, 2006, p. 265.
[88] Smith, 1982, p. 185.
[89] Schelling, 1978, p. 150.

satisfied". But A does not intend "to trigger a chain reaction through which complete segregation will be produced". A intends "to improve his situation by moving to a mildly segregated neighbour", but A does not intend to "participate to an 'unravelling' process"[90].

As these examples show, an invisible-hand explanation is made of two sorts of description. It first describes what agents do intentionally, taking the same perspective as the agents involved, by giving a description of their action that the agents would be able to offer, were they asked to do so. It secondly refers to these actions under a description that is unknown to their performers.

2. Known and unknown descriptions

So far we have seen that an invisible-hand explanation says what it is that agents are doing. First from their own perspective, and then from some other inaccessible perspective. The next question is: What exactly is the nature of the perspective from which agent's actions must be re-described? Most invisible-hand theorists seem to agree to roughly describe it as a *theoretical* perspective. For example, Rothschild sees in the gap between agents' un-theoretical point of view and the explainer's theoretical point of view, one central feature of invisible-hand explanations. "The invisible hand", she says, "presupposes the existence of a theorist (if not a reformer), who sees more than any ordinary individual can. The disembodied hand is invisible to its millions of petty subjects, but it is visible to 'us': to theorists"[91]. There are descriptions that agents who are involved in an invisible-hand explanation cannot act under. A description that is too theoretical will in particular not be appropriate. This is because agents led by the invisible hand are supposed to lack a theoretical stance on what is happening. A certain gap is required between what agents led by the invisible are aware of and what they are actually doing. The agents' commonsense point of view contrasts with the scientific skill that is required to ascribe any causal role to human propensities and actions in the social order of things.

There is however more than one way to act from a non-theoretical or commonsensical perspective. What is the theoretically limited point of view from which invisible-hand agents view their actions? What precisely are they blind to? What is it that theorists can see and that agents involved in an invisible-hand explanation are unaware of?

Interestingly, the examples we have just reviewed deliver different answers to these questions. They reflect the various ways a description can sound more theoretical than another. The theorist's description diverges from the agents' description according to the following multiple (but compatible) criteria:

First, agents might be unaware of the good they do to others. The egocentric perspective from which they act prevents them from seeing that, by acting selfishly, they are producing benefits for all. It is this difference in particular that many Eighteenth Century scholars pay attention to. Smith is famously eloquent on this matter. He accurately observes that agents do not let any other-regarding considerations constrain their pursuit of profit. He notes that agents do not describe

[90] Schelling, 1978, p. 150.
[91] Rothschild, 2001, p. 124.

what they do as helping the humanity but as fulfilling their *own* convenience with a view to showing how the former consequence can very well be added to the latter. Only the political theorist, however, sees how everyone ultimately benefits from the profit-motive of some.

Agents might alternatively be unaware of the *long-term* effect of their action, given their short-term perspective. Ferguson refers to this sort of weakness when he says that "every step and every movement of the multitude… are made with equal blindness to the future"[92]. The same ignorance of what comes next was noticed by Mandeville who believed that "the shortsighted vulgar in the chain of causes seldom can see further than one link"[93]. Whereas the myopic mind does not perceive how vices turn into virtue, "those who can enlarge their view, and will give themselves the Leisure of gazing on the prospect of concatenated events, may, in a hundred places see good spring up, and pullulated from Evil, as naturally as Chickens do from Eggs."[94]

Brennan and Pettit conceive the sort of blindness that characterizes agents involved in an invisible-hand process still differently. Agents who are led by an invisible hand, they say, distinctively "lack any sense of the aggregate shape of things":

> Those who remain mere participants in the system, those who fail to adopt a theoretical stance on what happens, will necessarily fail to recognize what is going on. It is definitional of not having attained a theoretical perspective on the system of distribution – of not having attained Smith's insights, in particular – that one does not recognize the operation or the effect of the sanctions in question. Participants who are not also theorists are embedded in their individual positions and are aware of the immediate pushes and pulls that work on them; but they lack any sense of the aggregate shape of things. The hand that moulds things to that shape remains invisible within their perspective[95].

The more an agent is able to see beyond the "immediate pushes and pulls that work on them"[96], the more the description he is able to offer of his action will sound theoretical. In other words, the more an agent is aware of the causal role of his action in the bringing about of an outcome, the more the way he describes his action will sound scientifically grounded.

Finally, agents could well be unaware of the scientific terms with which their actions could be re-described. Taking a folk-psychological perspective, they describe their choice in folk psychological terms. For example, they don't look for goods with "high saleability"; they look for goods that are "often exchanged". They do not assign a function of a "medium of exchange" on a certain good; they rather expect it to be more easily exchangeable on the marketplace than another. They do not believe that using a more exchangeable good will "reduce the transaction costs of the barter system". They rather believe that it will enable them to "acquire the good they want in a quicker way". Agents are "ordinary" in a sense that they use ordinary terms. Unlike the theorist who re-describes what agents do using scientific notions.

[92] Ferguson [1771], 1782, section II.
[93] Mandeville [1723], 1988, p. 123.
[94] Mandeville [1723], 1988, p. 123.
[95] Brennan & Pettit, 1993, p. 200.
[96] Brennan & Pettit, 1993, p. 200.

We have distinguished four ways of setting apart the agents' un-theoretical perspective from the explainer's theoretical point of view. We have also shown how within each of these possibilities, things vary by degrees. Agents can be more or less selfish, more or less aware of the aggregate shapes of things, relatively focusing on the short-term effect of their action and more or less familiar with scientific descriptions. In sum, the knowledge and motive of agents who are led by the invisible hand are not sharply defined but have vague - indeterminate - boundaries.

3. Stretching the sphere of intendedness: how far?

The distinction between what belongs to the sphere of intendedness and what stands outside its limit corresponds to a well-established distinction between a "result" and a "consequence"[97]. A result of an action is conceptually connected to the action that brings it about. It figures in the reason for which agents act and so it is not possible to describe the action without mentioning it. By contrast, a consequence of an action is externally or causally connected to that action. The final stage of an invisible-hand process is therefore a consequence of actions rather than a result.

By moving from one description to another we turn consequences into results and results into consequences[98]. Take, for example, the actions of the managers of the protective agency. The dividing line between the results and the consequences of their actions can be similarly relocated. Depending on the sort of explanation one intends to offer, it may be moved from the description "giving compensation to the independents" to the description "setting up a minimal state". Of course, if the latter description were the description under which the managers acted knowingly, it would not qualify as an invisible-hand *explanandum*. An intentional-design explanation of the rise of the minimal state would have been given. The point is that by moving from one description to another, we turn intended consequences into unintended ones and *vice versa.*

How far can the dividing line between results and consequences be shifted? It seems that the sphere of intendedness cannot be endlessly inclusive. Its volume encounters some limit, which, on the conventional conception of the invisible hand so far reconstructed, is generally expressed in terms of the agent's lack of theoretical knowledge. There is however no clear-cut boundary between ignorance and awareness. To be sure, there is a clear-cut conceptual distinction between ignorance and awareness. But their extension —some intermediate cases — will be in*determinate*. We therefore need to figure out what level of theoretical knowledge agents who are led by the invisible hand shall be (dis)allowed to have. The problem is that there is no straightforward answer to this question.

Cowen raises this problem when, in his discussion of the notion of unintended consequences, he asks: "how fine a description of the final outcome [...] agents [...] must intend in order to support a classification of the mechanism as an intended or unintended consequence?"[99] He takes as an example the mechanism by which segregation arises in the Schelling's checkerboard city and finds it impossible to

[97] Von Wright, 1963, pp. 39-41.
[98] Mäki, 1990a, p. 163.
[99] Cowen, 1997, p. 134.

appraise on the basis of the distinction between intended and unintended consequence. As he notes, there is no straightforward way to determine whether the resulting segregation is an intended or an unintended consequence:

> Does the answer depend upon whether some agents actively desire to created full segregation, whether agents know that their collective actions will lead to segregation, or whether agents are entirely unaware of the mechanism? How fine a description of the final outcome must agents intend in order to support a classification of the mechanism as an intended or unintended consequence?[100]

Our intuitions are, on these questions, more than likely to clash. The reason, I have argued, is the grey zone that stands between blindness and clairvoyance. There isn't any obvious theoretical threshold under which agents should be considered to be involved in an invisible-hand process. For this reason, relying on a gap between agents' blindness and the theorist's farsightedness as a way of sorting out invisible-hand explanations from intentional-design explanations leaves many cases unresolved. Stated as such, the distinction therefore does not convincingly play the classificatory role it is supposed to.

The existence of unclear cases, some will here object, does not show that there are not clear cases of unintended and intended consequences. Admitting a grey zone in the application of a classificatory criterion does not annul its effectiveness insofar as there are cases that plainly meet and do not meet it. My way of responding to this remark is to observe that even those uncontroversial cases of intended and unintended consequences will be harder to list than we may think. In my own experience, what some regard as clear cases of un/intended consequences, others do not. I have found it hard, for example, to show that those who consider the final stage of Schelling's checkerboard city game as somewhat satisfying its residents' initial preference for mildly segregated areas have a curious way of distinguishing the intended from the unintended. What difference does it really make, the sceptics would promptly retort, whether one finds oneself living in a fully segregated neighbourhood instead of a mildly integrated one? Isn't being prepared to tolerate the latter being satisfied with the former? Of course one could rightly reply to the sceptics that if someone wants to live in a mixed yet majority white neighbourhood, she clearly is not going to be satisfied with a fully segregated neighbourhood and that those who see no difference in whether satisfaction conditions have been met in both scenarios are just wrong. In order to resolve the debate, one will thus have to figure out whether living in a mixed neighbourhood is positively appraised (total segregation would in this case be an unintended consequence) or whether it is a situation that residents merely tolerate (total segregation would then enter agents' sphere of intendedness). In all cases, what the example shows is that it is not straightforward whether some situation is intended or not.

This indeterminacy explains some quarrels between invisible-hand advocates and their opponents. For example, the dispute between Hayek and Petsoulas[101] referred to in the introduction might, at least in part, be understood as a clash of intuitions regarding the degree of theoretical awareness agents involved in an invisible-hand process shall be allowed to possess. While Petsoulas believes that agents following Hayek's rules of conduct are not blind enough to be led by the

[100] Cowen, 1997, p. 135.
[101] Cf. Petsoulas (2001).

invisible hand, Hayek puts forward less demanding requirement as to what invisible-handed led agents cannot see.

Relying on a gap between awareness and blindness also has another undesirable implication. It makes the task of showing that a given outcome could be the result of an invisible-hand process too easy. On the conventional view, unintended consequences are obtained by finding a description of an agent's actions that are theoretically inaccessible. However, for any action, there will always be a description that is sufficiently abstract or sophisticated enough to have escaped the grasp of their performers. Vernon objects to the intended vs. unintended distinction along these lines. Even an apparently entirely intentional action like driving a nail into a piece of wood, he argues, can count as unintended. "For even if one drove the nail in at the precisely intended point to the precisely intended depth one would not predict or intend the precise arrangement of wood fibres brought about by the nail's entry"[102]. There does not seem any circumstance in which it makes sense to intend to drive a nail in some precise arrangement of wood fibres rather than in some others.

A few invisible-hand explanations seem to be obtained in this dubious manner. Consider, for example, Nozick's explanation of the minimal state. While there are different descriptions under which the managers of the dominant protective agency act, "playing one's part in the establishment of a minimal state" is, by Nozick's hypothesis, not among them. But could it ever be? Isn't referring to a social outcome as a minimal state to give it a description that is likely to have escaped the mind of those who contributed to its rise? As theoretically skilled as agents may be — that is, as acquainted as they may be with Nozick's first chapter of *Anarchy, State and Utopia* — they usually do not refer to the effect of what they do in such abstract and technical terms. In fact, there does not seem to be any circumstance in which it would make sense to intend to play one's part in creating a minimal state. Although Libyans may be described as trying to do just that, as some here reply, they certainly could not be ascribed such an intention. The reason is that, like most ordinary people, they do not have a clear concept of what a state is, i.e. a non-biological person which has duties and makes promises and enjoys sovereignty. Because it involves the unusual mastery of a theoretical concept, attempting to institute a minimal state is something one can only do unintentionally. One might think that even a Libyan tribesman typically intends to bring about "the sort of state they have in Northern Europe". But this description still falls short of constituting a grasp of the concept of a state.

Is there a more compelling way of construing the distinction between intended and unintended consequence to make it classificatorily more robust? One possibility is to consider the sphere of intendedness as being *conceptually* limited. On this new possibility, there are consequences of actions that will never be results and *vice versa*. In the first case, the invisible hand is the only conceptually available explanatory option. In the second case, appealing to the invisible hand is conceptually forbidden. These two situations are the topics of the next two sections.

[102] Vernon, 1979, p. 67.

4. Consequences that can never fall within the sphere of intendedness

How far can the sphere of intendedness extend? Quite far, according to Mäki, but not infinitely so. There are consequences, he argues, that will never be results. These consequences will never be realized in virtue of being intended, however theoretically sophisticated agents may be. Consider the example of A intentionally opening the window. According to Mäki, "it may be true that A intentionally cooled the room, but it is probably not true that it was an intended result of her act of cooling the room that she could not deliver the presidential address."[103]

The example, however, points to an improbable scenario and not an inconsistent one. What is described as "probably not true" remains conceptually possible. Being unable to deliver the address could, on some admittedly unusual situation, be an intended result of my opening the window. Panic-stricken by the prospect of giving a speech, I may do anything possible to avoid it, including making myself vulnerable to any debilitating flue by cooling down the room.

We are looking for examples of effects that, however far-sighted, clever or theoretically sophisticated agents might be, can *only* be derivative. To explore the conceptual limitation that prevents a certain consequence from ever becoming a result, I will consider (and expand on) Elster's account of the old idea that there are "states that are essentially by-products"[104].

4.1 States that are essentially by-products

States that are essentially by-products are, on Elster's characterization, "inaccessible states"[105] or "inherently out of reach for intentional action"[106]. These are states that can only occur as an effect of a behaviour that is directed toward some other end. Agents have to abstain from trying to bring them about for these states to come about. Spontaneity, indifference, sleep, being autonomous, love, being amused, being sexually aroused, forgetting, believing, self-respect, virtue, making an impression, being admired, are examples discussed by Elster. Any attempt to bring these states about is bound to fail. The issue is not whether it is hard, but whether such any procedure can be implemented at all.

Elster explains the failure as one that involves a contradiction. The contradiction is between the lack of intentionality that characterizes spontaneity, indifference, and so on, and the intentional component that always accompanies the intention to elicit a mental state either in oneself or in others. Because being spontaneous is to let go, the effort involved in trying to achieve it cannot but be self-defeating. On close scrutiny, however, there is more than one reason why a mental state eludes anyone trying to bring it about. Although Elster tends to mix them together, different cases can in particular be distinguished.

There is, first of all, the case illustrated by sleep. The attempt to sleep is doomed to fail, according to Elster, because "it requires a concentration of mind that

[103] Mäki, 1990a, p. 163.
[104] Elster, 1983a, pp. 43-108.
[105] Elster, 1983a, p. 43.
[106] Elster, 1983a, p. 56.

is incompatible with the absence of concentration one is trying to bring about"[107]. Attempting to be natural is similarly contradictory. He who tries to be natural will have the impossible task of somehow "not to give the impression that [he] is trying to make an impression"[108]. Another example is indifference. As Elster explains, "the intentional element involved in the desire to appear indifferent is incompatible with the lack of intentionality that characterizes indifference"[109] (note that the lack of intentionality here has the non-technical sense of lack of effort and not the philosophical sense of lack of aboutness). Indifference remains what philosophers call an intentional mental state in the sense that there is always an object toward which one is indifferent. But to be indifferent it not something one can intend in so far as, according to Elster, the effort involved in the intention to be so is not compatible to the lack of attention that characterizes it[110]. Essentially by-product states of this first type are "privative states". They involve "the absence of a specific form of consciousness such as attention to the impression one is making, or the absence of consciousness in general"[111]. Trying to bring about these states is trying to repress a thought. One does not fall asleep when the mind is encumbered with thoughts (such as the thought that if one does not fall asleep tomorrow will be a disaster). The desire to fall asleep can only be counterproductive because it involves a certain form of attention and that "attention — unlike valves — cannot be shut off at will"[112]. In other words, these are mental states that one cannot intentionally elicit in oneself because eliciting something involves effort and lack of effort is what characterizes these states. Elster also uses a distinction between "external negation" and "internal negation" to explain why certain states cannot be willed. To be genuinely indifferent, I must *not* will to be unconcerned. To have the will to be *un*concerned is already to care too much.

Elster finally suggests how these states could nonetheless be intentionally elicited. Being indifferent or being spontaneous, he argues, are character traits that could be acquired by "becoming a certain kind of person — a person who could not care less about making an impression"[113]. To what extent being a certain person is something that, in turn, can be willed is a difficult question[114].

Cases of essentially by-product states are not always of that first type. Mental states that are "instrumentally useful, and yet out of reach for instrumental rationality"[115] fall into a different sub-category. The belief that god exists is an example. The utility of holding it is not a reason to adopt it. In particular, the pleasure, the serenity, the self-boost in self-esteem or whatever desirable effects of

[107] Elster, 1983a, p. 45.
[108] Elster, 1983a, p. 45.
[109] Elster, 1983a, p. 45.
[110] As Otto Bruun notices, however, there are lots of things one can intend that (i) require no effort or (ii) precisely involve a cessation of effort: the intention to slack off, the intention to give in to an urge, etc. In order to highlight the problem, the distinction between a "prior intention" (i.e., the plan that one has before the performance of an action) and an "intention-in-action" (the intention one has during the performance of the action itself (Searle, 1983, pp. 83-98) may be useful. In light of such distinction, the intention to be spontaneous may turn out to be not so contradictory as the effort involved may be a feature of the prior intention to be spontaneous rather than a feature of the intention-in-action.
[111] Elster, 1983a, p. 46.
[112] Elster, 1983a, p. 48.
[113] Elster, 1983a, p. 45.
[114] As Elster recalls, Pascal offers some clue as how it might well be. On their view, acting as if one were in love, courageous, virtuous or religious will ultimately make us so.
[115] Elster, 1983a, p. 51.

having a certain belief cannot be a reason for holding it. Only evidence impartially gathered in favour of such a belief is. Having a belief is therefore essentially a by-product of the activity of pondering the evidence supporting it. To be sure, pondering the evidence is an intentional activity. But coming to the conclusion that the evidence on balance supports the belief is not.

Why are beliefs in a separate sub-class? Unlike indifference or any mental states of the first type, having a belief is not a privative state. Holding a belief does not involve the absence of consciousness or the lack of attention that characteristically defines unaffectedness, being natural or being indifferent. The reason why a belief escapes any attempt to intentionally have it is that it is the end state of a chain of previous mental states, none of which can be intentionally elicited in the first place.

Being admired will be the last case to consider. According to Elster, being admired is an effect that can be brought about "knowingly and intelligently", albeit not intentionally. Elster is not as clear as one might wish about the difference between these two modes. He says that being admired cannot be one's *sole* or even *main* goal[116], implying that it could very well be one's lesser goal. So having the intention to behave admirably must prevail over having the intention to elicit admiration. In some other passages, however, Elster is much more restrictive regarding the extent to which being admired could be willed. Admiration, he says, is not felt toward agents who seek it. If this is true, admirable persons are required not to care about being admired. As Elster also says, "nothing is so unimpressive as behaviour designed to impress"[117]. If your goal in doing something is to be admired you will end up behaving in a way that is not admirable. Surely, we all know that if we do something admirable, admiration will accrue to us. Still, being admired cannot, according to Elster, be our purpose when we are doing something admirable[118]. The purpose of our action must lie elsewhere, like getting it right. Elster's brevity leaves us guessing why this should be so. One explanation is that being too preoccupied by my image prevents me from performing the admirable action well. This in turn can be because I do too many things at once, as when, too concerned by the effect of my performance on the audience, I lack the required concentration to achieve anything worthy. Another explanation is that being preoccupied by the other's approval is itself sufficiently contemptible to cancel out all the virtual goodness of one's admirable action. Be that as it may, Elster will perhaps admit that once having obtained the other's admiration, you can freely pay attention to the admiration produced without impeding it. The admirer may even like it when the object of his affections shows awareness of the admiration. This way, she expresses her gratitude to her admirer.

To conclude, there is more than one reason why some states cannot be willed. Sleep is an essentially by-product state in virtue of excluding the mental effort that is always involved in the attempt to obtain anything. Belief is an essentially by-product state in virtue of being a state one finds oneself in as the consequence of other states that were no more deliberately induced. Being admired is an essentially by-product state in virtue of the disapproval induced by any attempt to be admired. These three

[116] Elster, 1983a, p. 57.
[117] Elster, 1983a, p. 66.
[118] Elster, 1983a, pp. 55-56.

sub-types of essentially by-products states may not cover all the examples provided by Elster[119]. They however will suffice for present purposes.

4.2 The only conceptually available explanatory option

We have seen how some mental states can escape those who try to reach them. The fact is however not sufficiently acknowledged, according to Elster, who calls the "intellectual fallacy of by-products" the failure to recognize it. The fallacy is defined as the attempt "to explain [an essentially by-product state] as the result of action designed to bring it about"[120]. The fallacy mainly prevails in social science, according to Elster, where attempts to explain essentially by-products by reference to an agent's intention to bring it about abound. In other words, we may say that the fallacy occurs when an intentional-design explanation is advanced where an invisible-hand explanation is the only conceptually available option.

The fallacy occurs in particular when we misconstrue effects that can only be by-products as political goals. Elster gives, as an example, the psychological benefits that are derived from political participation. Being enrolled as a juror, adopting democracy or fighting for a political cause, it is rightly observed, will bring to the participants some educational effects. It will in particular confer on them self-respect, class-consciousness, and an opportunity for self-realization. What is fallacious is to further claim that those useful effects can very well be the main or even the unique reason to engage in political participation. The main goal of democracy, of the jury system and of political participation, it is argued, lies in their ability to educate the citizens, the jurors or the militants as well as to induce in each of them a sense of self-respect. The intellectual fallacy of by-products consists in claiming that even if militants have no impact on policy outcome, political participation remains justified as an occasion for gaining such self-respect.

Elster grants that self-respect can undoubtedly result from all kinds of political commitments. Yet it is a benefit that only those who believe that they are doing something that has value beyond their personal development can collect. One cannot acquire self-respect unless one has disinterested reason to do so. Self-respect depends exclusively on my belief about how much I have accomplished. I need to be assured that, as a political participant, I am not merely playing a game. To be sure, there are people defending vain causes. But they have to fool themselves, by wrongly believing they are making progress in advancing their cause. No one can take part in political advocacy if his only reason is to build up one's self-respect.

Consider the circumstances of agents when they collect the psychological rewards of some political activities. If Elster is right, explaining this situation as the intended results of the agents' decision to be political activists would be mistaken and the mistake would be of the conceptual kind rather than of the empirical one. Being essentially a by-product, the situation cannot be accounted for other than by an invisible-hand type of explanation. Appealing to the invisible hand is not, in this case,

[119] Chan & Miller (1991) mentions a fourth case, unnoticed by Elster, where "the desired outcome arises in a chance manner without the agent being able to increase the probability of its arising by intending it". They give the example of a poor dart player who "hit the bull's eye or simply throws his dart in the general direction of the board. Here hitting the bull's-eye is a by-product of throwing at the board: aiming at the bull's-eye neither increases nor decreases the chance of arriving at the desired outcome" (Chan & Miller, 1991, p. 98).

[120] Elster, 1983a, p. 43.

an alternative way of explaining the social reality. It is the only conceptually available explanatory option.

I have so far attempted to circumscribe the class of social phenomena that can only be explained in invisible hand terms. Are there any other cases falling in that class? Reflecting on this question, some have noticed that any individual agent may intend to F and *hope* that her F-ing will be part of a macro-social process. She can even *predict* (rightly or wrongly) that the macro-process in question will come about. However, she cannot *intend* her F-ing to be part of such a process since its being part of such a process is essentially not something directly subject to her will. Being part of a macro-social process is something that I cannot will, let alone intend to bring about the macro-social process of which my action is a part. It follows that all macro-social states can only be invisible-handedly explained inasmuch as being part of them is something that each agent cannot intend. Taking this argument at face value, all that a social outcome needs to be eligible for an invisible-hand explanation is to be a macro-social outcome, such as a revolution, a war or a coup. Yet numerous intentional-design explanations have been given of the latter. Shall we conclude that they are all wrong?

Fortunately, this conclusion can be evaded. To see how, take, as an example of such a macro-social state, a forthcoming revolution. I, as a future participant in that revolution, can very well intend to protest in the street. But according to the above argument I cannot intend that my protest be part of the revolution. The process of which my action is a part is something that I cannot intend for to happen in the first place. Such a process is not subject to my will because the only thing that is subject to my will is my own participation in it. Why, we may ask, can I not have the intention that a process happens? The reason, still yet to be expounded, seems to be the following one: I cannot have the intention that the others protest because intentions require, as a condition of satisfaction, that the realization of their content — that what is intended — is under its bearer's control[121]. Others' actions are not under my control. The grammatical awkwardness of formulae such as "I intend that you do your part" is revealing. Large-scale processes are, in sum, essentially unintended consequences in virtue of the impossibility for any of their participants to intend that their co-participants play their part.

The way to avoid this conclusion is to observe that although I cannot intend that you will do your part of the revolution, I can perfectly have some effect on your behaviour by convincing you and many others to protest. I can, for example, draw your attention to the social injustices our current leader has produced in order to bring about some rebellious impulse in you. In light of such indirect influence, I may after all perfectly have the development of a process as the content of my intention. And this conclusion in turn refutes the suggestion to extent to all macro-social processes the class of social phenomena that can only be invisible-handedly explained. To recapitulate, revolution or any large-scale social phenomenon can be, at least indirectly, subject to anyone's will inasmuch as anyone can persuade the others to be part of it. If revolution can be intentionally launched, being part of it is something anyone can, at least indirectly, intend to.

Outcomes that can only be invisible-handedly explained may be harder to find than it may seem.

[121] Velleman, 1997.

5. Consequences that can only fall within the sphere of intendedness

In the previous section, the goal was to investigate whether the invisible hand has a privileged domain of application. This time, the goal is to examine whether intentional-design explanations also have an exclusive scope. Is there a class of outcomes that can only be produced intentionally?

Rothschild alludes to this question when she considers whether there could be certain domain of activities for which invisible-hand explanations may "work less well". Activities that involving "reflective thinking"[122], she claims, cannot be performed by agents who are led by the invisible hand. Conversely, invisible-hand explanations "work best when the activities in question (such as bartering one sort of commodity for another) are relatively unreflective, and relatively unencumbered with social theorizing"[123]. Because economic activities are often highly reflective[124], she further claims, invisible-hand explanations of economic life are often unconvincing.

Unfortunately, Rothschild leaves one to guess under what conditions an action is "unreflective" or "unencumbered with social theorizing". The notion of "pure action", developed by Moya, is, in this respect, illuminating[125]. Pure actions are, on Moya's definition, actions that cannot be non-intentionally performed. Cases in point are greeting, signalling for a turn, or marrying. As Moya observes, "there is no such thing as greeting, signalling for a turn, or marrying unintentionally. To do it intentionally is a necessary condition of greeting, signalling for a turn, or marrying"[126]. By contrast, shooting a gun, killing someone or scoring a goal is not necessarily intentional. As Moya says, to be able to attribute the action of greeting to someone, she has to perform that action under the description "greeting". The same is true of marrying. "'Marrying' cannot be a true description of what someone does if 'marrying intentionally' is not"[127].

Pure actions, according to Moya, can be distinguished from their non-pure counterparts in the following manner. First, pure actions are not subject to mistakes. I cannot mistakenly greet someone (as opposed to being mistaken about the person I am greeting). Surely, some movements of mine can be mistakenly taken as such. But these movements are not a greeting if I do not intentionally make one. Note that this feature also applies to the act of bidding at an auction. While there is such a thing as accidentally shooting someone, scoring a goal or breaking a glass, there is no such thing as bidding at an auction without meaning to. While one can break a glass "by mistake", one does not find oneself buying an object unless one has it as one's purpose.

Second, we recognize actions that are pure by the fact that the actor cannot discover or know observationally that she is doing them. I can discover that I am absent-mindedly raising my arm while attending an auction. I cannot discover that I am signalling a desire to purchase the painting at a higher price. The unlucky shooter can discover that he killed someone because his action was not part of his intention,

[122] Rothshild, 2001, p. 142.
[123] Rothshild, 2001, p. 142.
[124] Rothschild, 2001, p. 142.
[125] Moya, 1990, chap. 4, 5.
[126] Moya, 1990, p. 52.
[127] Moya, 1990, p. 53.

which was to clean up the gun. There is, however, no accidental seller, no accidental buyer, and no accidental auction bidder.

Thirdly, whereas non-pure actions involve happenings, pure actions are pure in virtue of the fact that no happening is essential for their performance. Scoring a goal, for example, essentially involves the happening that the ball goes through the goal-posts. The same is true of the action of killing someone. Killing someone involves essentially the happening that someone dies.

Happenings can occur as the result of the performance of action *or* as the result of the occurrence of another happening. Consider, for example, the action of raising one's arm and its happening that one's arm is rising. For this happening takes place, we do not need the action of raising one's arm. One's arm rising is a happening that can occur with nobody raising their arm. One's arm can rise simply by being raised by a machine. By contrast, pure actions have results — rather than happenings — that cannot occur without the action of which they are results being performed[128]. A greeting can only occur as the result of the action of greeting. A greeting is therefore not a happening.

Now it is a feature of actions that essentially involve a happening, that they can be performed without the agents acting intentionally. Someone can score a goal unintentionally as when, kicking the ball a player back-handedly directs it into his own goal. There are however actions that do not essentially involve happenings and these are cases of "pure actions". Examples are signalling for a turn, making an offer, marrying and holding a lecture. Trying to separate, for these kinds of actions, the action performed from its happening, is impossible. This is because, as Moya notes, there is no real difference of content between these two. For example, there is no difference between "Someone made an offer" and "An offer took place". Similarly, there is no real difference of content between "Someone held a lecture" and "A lecture took place". By contrast, there is a difference of content between "Someone killed Smith" and "Smith's death took place". Or between "Someone raised his arm" and "An arm's rising took place". Another way to grasp the difference between pure and impure actions is to say that, for pure actions, the results are actions not happenings[129]. Actions are pure in the sense that their performance is all there is.

The reason why pure actions are always intentional relates to their having no happenings as results. As Moya says, "if no specific happening can be brought about, no specific happening can be brought about in an unintended way"[130]. The argument seems to be the following. It is when there is a gap between an action and its happening that a lack of intentionality can find room to exist. Being essentially associated with no happening, pure action leaves no room for lack of intentionality.

Now that we know what pure actions are we can see why they may on occasion play a crucial role in disqualifying the relevance of an invisible-hand explanation. If a social state of affair consisted entirely of pure actions, it would certainly inherit the main feature of these pure actions, namely their intentionality. Marriages are necessarily intentional to the extent that it consists in various actions all of which requires at least awareness on the parts of their performers. Ruben "no one could get married in a society in which everyone was unaware of the fact that anyone ever got married at all"[131]. An auction is another example. No one could bid at an

[128] Moya, 1990, p. 38.
[129] Moya, 1990, p. 38.
[130] Moya, 1990, p. 65.
[131] Ruben, 1998, p. 438.

auction in a society in which everyone was unaware of the fact that anyone was bidding at an auction. It entails that the situation of agents when they bid at an auction is a situation that cannot be construed as the unintended consequence of agents' action. These states of affairs, together with marriages, will never be eligible for an invisible-hand explanation. The point of this section is to show that the reason for this ineligibility is conceptual rather than practical. It does not depend on some contingent lack of sense of the aggregated shapes of things on the part of the agents involved —it does depend on something that could be otherwise. It rather depends on the conceptual impossibility of construing the situation as an unintended consequence[132].

Conversely, invisible-hand explanations should refer only to impure actions. This is because only impure actions can accommodate the lack of intentionality that is essential to all actions in virtue of which the outcome to be explained in invisible hand terms is obtained. The claim will however sound problematic to those who, like Searle[133], defend an intentionalist account of institutional reality. On such account, institutional facts are language dependent[134]. They are created by linguistic acts and in particular by speech acts and not by any other sort of action[135]. Speech acts, according to Searle, are solely responsible for the creation of institutional facts. There is no social fact, Searle argues in particular, that is not essentially based on a declaration that P. He then defines a declaration as a speech act that "change[s] the world by declaring that a state of affairs exists and thus bringing that state of affairs into existence"[136]. To be declared married is essential to be married, to be declared a leader is essential to being a leader and to be declared a university is essential to being one. Other speech acts such as commanding, promising, accepting, recognizing, requesting, pronouncing, decreeing, etc., that P also play a crucial role, according to Searle, in the creation of institutional facts.

Speech acts — promising, ordering, baptising, marrying, etc. — are all pure acts in Moya's sense. It is impossible in the case of a speech act to hive off the happening — e.g. Catherine and Jules's marriage is being pronounced — from the act itself — e.g. someone pronounces Catherine and Jules to be married. And one does not find oneself declaring them married unless one intends to do so. But if it is true that speech acts are essential for all social facts, either because they are essential to bringing them about or because they are essential to keeping them in existence,

[132] This conclusion might however be shaded in the following way. We have been speaking so far of *social situations* rather than *social patterns*. These two notions relate to one another in the following way. A social pattern is made of recurring social situations (or social happenings) of the same kind. For example, marriages, auctions, sells, are social situations. But an increase, a decrease or stability in rates of marriage is a social pattern. We have shown that an invisible-hand explanation cannot account for the occurrence of a social situation that involves the performance of pure actions. Consider however this social situation as recurring in a steady (or in an increasing or in a decreasing) fashion, and you will have obtained a social pattern to which an invisible-hand may very well apply. Here is why. Although marriages cannot happen unless agents act under the description "I am getting married", a decrease of marriage can very well happen unbeknownst to all married persons. This is because the awareness that is essentially involved in a marriage as a social situation is restricted to *that* marriage and does not have to dwell on any of the other marriages taking place within a certain span of time. There is therefore no contradiction in offering an invisible-hand explanation of a falling rate of marriages even if no marriage can take place unbeknownst to the participants.

[133] Searle, 1995, 2010.

[134] For a more detailed and critical treatment of the reasons why institutional facts are language dependent, see appendix, p. 127.

[135] Prior statements of Searle's social ontology as one involving speech acts can be found in Reinach (1913, Cf. Mulligan, 1987 for a presentation of Reinach's theory of speech acts).

[136] Searle, 2010, p. 12.

they strikingly rule out the legitimacy of all invisible-hand explanations. The latter are only relevant in case a social outcome is obtained by virtue of impure actions, actions in which intentionality can be lacking.

Conclusion

The goal of this chapter was to answer the following question: are invisible-hand explanations persuasively distinguished from their rivals when their scope is delineated in terms of unintended consequences? The diverse conclusions we have reached in this regard can be summarized as follows.

Characterizing invisible-hand consequences as unintended consequences can be understood in two ways. On the first (standard) reading, a consequence is unintended in virtue of a lack of availability of concepts on the part of agents who have contributed to its occurrence. I have shown that this way of construing the notion of unintended consequences leaves most cases undetermined. The reason is that, I have argued, we have no clear-cut and intuitive idea about what concepts agents involved in an invisible-hand process are allowed to master[137].

On the second characterization, a consequence is unintended in virtue of some conceptual limitation. While some effects will always be side effects, others will always fall within the agent's sphere of unintendedness. I have dealt with the first with the help of of Elster's idea of social *states that are essentially by-products*. Moya's notion of *pure actions* was helpful in defining the second.

Is this alternative way of interpreting the intended *vs*. unintended consequences distinction more compelling? Under the related reading, assessing whether a given consequence really is an unintended consequence is no longer a matter of having to decide whether agents have a sense of the larger scheme of things or not. It therefore avoids the problems associated with that question. But it is also useless in regard to social consequences that are neither essentially unintended nor intended. Most paradigm cases of so-called invisible-hand consequences such as the emergence of money, the rise of the minimal state, residential segregation, are neither states that are essentially by-products, nor states that are obtained in virtue of agents performing pure actions. Whether these social patterns really are eligible for an invisible-hand explanation is currently debated. Restricting the scope of the invisible hand to states that are essentially unintended is of no help in deciding who, in these discussions, has the right classificatory view.

[137] And, correlatively, what concepts they are supposed to lack.

II. From lack of awareness to lack of we-ness

Introduction

In his article on "Collective intentions and actions" (1990), Searle gives the example of a group of businessmen who are all familiar with Adam Smith's theory of the invisible hand. They have learned in business school that the best way for them to help humanity is by pursuing their selfish interest and they consequently each form an intention to this effect. That is: each intends to do his part toward helping humanity by pursuing his own selfish interest and not cooperating with anybody.

Searle intends this example to show the following. Although they pursue the same goal, are mutually aware of each other's intention, and have mutually dependent intentions, still the businessmen are not acting together. Helping humanity is something they individually rather than jointly intend to do. The example supports the view that a we-intention is not the same as the sum of I-intentions. It is a primitive notion, as Searle says, that is, one that cannot be understood in terms of individual intentions, however intricate the latter might be.

From our perspective, the example also deserves attention for another reason. It construes the invisible hand as an anti-We-ness theory. The theory is presented as one that considers the capacity to act jointly as a superfluous component of the way the social world works. It is a theory that claims that wealth could be redistributed just as well, even if the wealthiest had no inclination to act together to this effect — even if they were not inclined to ask themselves "what shall *we* do to help humanity?" rather than "what shall *I* do?".

This chapter expands on the second idea, that of defining the invisible hand as an anti-We-ness theory, relying on the first idea, that of taking We-intentions as a primitive notion. I will show that approaching the invisible hand in this manner entails a deep revision of the standard account. It entails in particular a rejection of the idea that the invisible hand is about the unintended consequences of actions that are directed toward other ends. I will also show that the invisible hand is an anti-We-ness theory *only if* collective intentionality is a primitive notion. In other words, agents who are led by the invisible hand are not allowed to act jointly *provided* and to the extent that joint actions are different from an aggregation of individual actions.

1. The standard account and its problems

Let us start with an example. Consider again the phenomenon of residential segregation. One possible explanation is to relate it to someone's decision to establish it. Urban segregation is a social pattern whose occurrence it is natural to conceive in terms of political schemes. We are inclined to think that residents cannot be divided according to their social, cultural or racial status unless some designing mind devises a plan to bring this about.

Let us investigate this hypothesis further by asking who could have designed the segregation of cities into Blacks and Whites, poor and rich, Italians, Irish and

Chinese, etc.? Two possibilities come to mind. Residential segregation might first be the project of a malevolent legislator[138]. The Warsaw ghetto and apartheid are historical examples of this. Segregation need not be planned by a central authority to take place, however. Residents of the same cultural, social or racial class may find it pleasant or merely convenient to live together in the same areas. Most Chinatowns in the USA result from its inhabitants' preferences for living in the same area.

Schelling offers an alternative explanation, as we have seen[139]. Residential segregation, he shows, need not be designed to take place. It can very well be the result of a preference for living in a mixed neighbourhood. His checkerboard city investigates how residential segregation could be the consequence of mild segregationist preferences among the residents. For this purpose, Schelling invites us to play a game[140] using a checkerboard, a few pennies and dimes in equal number and randomly distributed on the checkerboard. He then stipulates a few rules governing the moves of the pawns on the board: every dime wants at least half its neighbours to be dimes, and every penny wants a third of its neighbours to be pennies. Each can only move to the nearest empty square that meets their wants and they keep moving until they get satisfied. In the end of the process, the pennies and dimes form a segregated pattern. Here is an explanation that, unlike the explanations that naturally suggest themselves, does not describe residential segregation as being designed. Neither the residents, nor their rulers planed it.

If designers had been involved, residential segregation would have been an intended outcome. It would have been what a central authority or the residents themselves intended to bring about. By contrast, Schelling's explanation describes the separation of Blacks and Whites as the unintended consequence of their individual decision to move to mildly integrated neighbourhoods.

This is, by all accounts, what invisible-hand explanations do. Invisible-hand explanations are widely defined as being about unintended consequences of actions directed toward other ends[141]. This type of explanation distinctively accounts for some collective outcome in terms of the actions of many individual agents, none of whom intended to produce that outcome in particular but who interacted in such a way that the outcome arose nonetheless. Unintended consequences thus are what invisible-hand explanations are about. They define their scope.

This way of characterizing the invisible hand raises some difficulties, however. First of all, defining the invisible hand in terms of unintended consequences presupposes a certain degree of blindness on the part of agents that is close to idiocy. Often the relation between what agents individually do and what they, as a group, bring about, is so obvious that their lack of awareness starts to look like sheer stupidity. Zaibert raises this point in regard to Nozick's explanation of the emergence of the minimal state. "Unless one assumes that people in the state of nature are actually stupid", he argues, "there is no reason to suspect that they will not be aware that by doing so and so they are in effect, or might be, creating a state…Thus, in many instances of invisible hand explanations, people would have the explained

[138] Legislators are also referred to as the "state", the "government", the "central authority", the "iron hand", the "social planner" or as "the Great Men".
[139] Schelling, 1978.
[140] Schelling, 1978, pp. 147-166.
[141] Nozick, 1974. Ullmann-Margalit, 1978, Mäki, 1990a, b, Aydinonat, 2008.

phenomenon "in mind," even though not necessarily in the form of an intention, and even though they might not be bringing the phenomenon about intentionally"[142].

Identifying the second problem was the goal of the previous chapter. As has been shown, while invisible-hand explanations are commonly defined as being about unintended consequences, paradigm cases of invisible-hand explanations have been criticized for not convincingly meeting this constraint. Many of them do not persuasively show that the outcome they explain really is an unintended consequence of actions. The disagreement, I have argued, is nearly impossible to solve. This is because intuitions diverge about how far from agents' view the consequence must be situated to count as unintended rather than as intended. Defining an invisible-hand consequence as an unintended consequence inconveniently leaves most explanations in social sciences unqualified.

We therefore need to reconsider the standard view about the invisible hand. We need to examine whether agents cannot have the social outcome of their action in mind if they are to be led by the invisible hand. We need to investigate to what extent the outcome can enter their "sphere of intention"[143] without ruling out the intervention of an invisible hand.

I will address this question by submitting the standard account to various revisions of roughly increasing force until a clearly inappropriate case for the invisible hand is encountered. As we shall see, quite a few revisions will be tested before such a boundary case is found.

2. To have the overall outcome in mind

On the standard account, invisible-hand agents are required not to intend the outcome to be explained. The dividing line between intended and unintended consequences is not easy to draw. Should it be forbidden for agents to have the intention to bring about a consequence for the latter to qualify as unintended? And if some designers are allowed to enter the group, what if any is the limit to the number admissible for the consequences to remain unintended?

Invisible hand theorists do not trouble themselves with such puzzles. They get around them by each time endorsing the most restrictive option. They first stipulate that *none* of the individuals involved in an invisible-hand process shall be allowed to act as a social designer[144]. No hidden social contractors or influential agents knowing better than the multitude, and directing the choices of the latter accordingly, are supposed to be operative. Second, the lack of intention to bring about the outcome that is to be explained turns out to be a bare lack of awareness. In Nozick's words, agents involved in an invisible hand process do "not have the overall pattern in mind"[145] while acting. Ullmann-Margalit also takes the same restrictive view. She says that agents cannot be aware and hence foresee the overall effects of their actions[146].

[142] Zaibert, 2004, p. 121.
[143] Mäki, 1990a.
[144] Nozick is in this respect an exception when he admits the possibility of impure cases in which some agents, among the participants, act with the intention to bring about the pattern that is to be explained (Nozick, 1974, p. 352, note 7). The invisible hand might accordingly be refined as admitting of degrees.
[145] Nozick, 1974.
[146] Ullmann-Margalit, 1978, p. 271.

On Tuomela's view, as well, the unintended nature of the pattern to be explained amounts to a lack of consciousness on the part of agents about what they are doing. Invisible-hand explanations, he says, involve "many individuals who are supposed to be minding (only) their own business unaware of and hence not intending to bring about the ultimate overall outcome."[147] The standard account can be summarized as follows:

SA: An IHE is intended to explain a consequence of many actions performed by agents *who do not have it in mind.*

SA prompts the following question: What if agents had the final outcome of their actions in mind while acting? There is a reason not to exclude this possibility right away: it would be unnecessarily restrictive[148]. Take the invisible-hand explanation Millar offers of the end of the slavery system[149]. The slave owners can very well recognize that their decision to hire their slaves will result in the end of slavery. Yet, to be causally efficient, such recognition would have to be the reason for which they chose to hire them. Doctors, to take a case in point, may act in a way which they know will lead to the death of a patient without intending to kill the patient. Being aware of the effect of one's action is one thing. Being motivated by the prospect of such an outcome is another[150]. An invisible-hand theorist should only worry about the latter case. Supposing that agents have the concepts needed to describe relevant consequences, someone advancing an invisible-hand explanation should only rule out the possibility that agents act in full knowledge of these consequences without intending to bring about these consequences. Her invisible-hand explanation will remain as parsimonious as before (if this is what she attaches most importance to) even if she lets agents be aware of what their actions amount to. An advocate of parsimony in explanation should only minimize the number of causal relations that are necessary to explain how the world could work. Merely being aware of what one's actions will bring about is an epiphenomenon. An explanation that accommodates such awareness is therefore as parsimonious as one that rules it out.

[147] Tuomela, 1984, p. 451. Menger (1985, p. 133) distinguishes between institutions that are the result of a common will directed toward their establishment (agreement, positive legislation, etc.) and others which are the unintended result of human efforts aimed at attaining essentially individual goals".

[148] Also, in ordinary language, a consequence does not have to be unforeseen to be described as "unintended". Take the situation in which everyone is asked to reduce his water consumption and most choose not to follow the directive, correctly anticipating that everyone else will similarly free ride. Although all correctly foresee the drought, no one can be said to have the intention to bring it about.

[149] Presented in details pp. 7-13.

[150] The same point has been advanced in the context of a discussion about the ingredients of functional explanations. Against Elster (1978) who argues that, in functional explanation, the effect that (retroactively) explains the social phenomena, must be *unrecognised* by the actors who bring it about, Grimen (1994) observes that "neither knowledge nor recognition is on its own causally efficient in a relevant sense for functional explanations of the maintenance of patterns of behaviour or institutions. The fact that an actor knows or recognized that *p* does not alone bring about anything which could invalidate a functional explanation on the one hand, or contribute to its explanatory force on the other. Hence there is no need to protect against the actors' (…) recognition of knowledge per se of the beneficial consequences of patterns of behavior or institutions. In order to be causally efficient, knowledge or recognition must be acted on or used, that is, it must enter among the actors' reasons for acting" (Grimen, 1994, p. 119).

Take, as another example, Menger's explanation of money. Acquiring cowry shells with a view to using them as a medium of exchange may be accompanied by the recognition that doing so will transform cowry shells into money in the group. To be causally efficient, such recognition must be part of the reason for which they choose cowry shells rather than coconuts. Only if agents choose cowry shells with a view to increasing their popularity for the ultimate purpose of establishing them as a universal medium of exchange, would the recognition be causally efficient. And only then, it seems, would the situation fail to qualify as the work of the invisible hand.

The point is that whether agents have the invisible-hand consequence of their action in mind or not while acting does not matter. What does matter is that such recognition does not play any role in their decision to act as they do. Agents shall *only* be protected against having the outcome figuring among their reasons for acting[151] and the standard account shall be revised accordingly:

R(evised) A(ccount)$_1$: An IHE is intended to explain a consequence of many actions performed by agents who may have it in mind, but not in the form of a reason for action.

3. Wayward causal chains

What constraints must an agent meet if he is to be led by the invisible hand? So far, we have reached the following conclusion. Being aware of the consequences of one's action, while remaining unmoved by their prospect is not disqualifying. Suppose, however, that the recognition of the outcome enters into the reason for which agents act. Would it preclude the operation of the invisible hand? It seems so. There is apparently no way we can avoid describing the outcome as being intentionally brought about. Yet the possibility that agents acted unintentionally, giving space for the working of an invisible hand, should not be excluded too quickly.

Consider, for example, the explanation Marx offers of the wealth of European countries in modern time[152]. These nations, as Marx recalls, were under the influence

[151] There might be a case, however, for not being so liberal. How could agents resist having the intention to end the slavery system, it could be replied, once they recognise it as a cheaply obtained effect of their initially selfish actions? It can however be replied that making sure that the end of the slavery system is unanticipated is just a guarantee that its recognition won't play any motivating role. Lack of awareness is in this regard important in so far as it ensures the behaviour is not intended in order to produce the effect to be explained.

[152] Cf. Elster, 1985a, pp. 22-24. The mechanism at play in Marx's explanation is similar to the one exemplified in La Fontaine's fable "The labourer and his children". The fable tells the story of three sons who were too lazy to work in the fields, as their father wished them to. The latter told them that there was a treasure buried in the grounds. Eager to get rich in a hurry, they overturned the soil in an unsuccessful search for the treasure, and in doing so made it so fertile that they did indeed get rich, although not in the way they had planned. Although the sons intended to get rich and ultimately got rich, they cannot be said to have intentionally gotten rich. Getting rich would have been something that they intentionally did if the sons had gotten rich by way of finding the treasure they were looking for. But wealth came to them through a very different process. In both Lafontaine's Fable and Marx's explanation, the recognition of the effect (getting rich) clearly is central to the reasons why agents act the way they do. But this effect plays its causal role in an unexpected manner. In both examples, what occurs corresponds to what was expected to occur. But the mechanism that is responsible of the

of mercantilism, the view that prosperity depends on the possession of gold. And yet, Mercantilism is, as it turns out a faulty theory: an economically prosperous nation is not one that possesses a lot of precious metals. It is however a theory that did bring wealth to the nations who endorsed it. It brought them wealth by motivating them to engage in efforts, e.g. to colonize new countries, to the proliferation of new needs, of new exchanges and of new commercial opportunities, that ultimately brought them real wealth. The mercantile system thus may have played a crucial, albeit unexpected role in the development of wealth.

What the example shows is that acting with the intention to bring about X is not yet a sufficient condition for doing X intentionally. The content of one's intention, e.g. to bring about wealth, may represent the state of affairs in which that intention is satisfied. Yet it does not ensure that the state of affairs has been intentionally brought about. This is because the identity between the goal and the state of affairs produced might be a twist of fate. Marx's explanation involves, I suggest, a wayward causal chain. It is important to note that what makes it a relevant case of the latter is not the fact that wealth was produced by the application of a false theory rather than by a good one. Had mercantilism been a sound economic theory, a wayward causal chain would still have been involved. This is because the wealth of these nations did not arise in conformity with mercantilist principles (whatever their worth), but rather as the side effect of measures taken in order to apply them. Although wealth happens to be what mercantilism pursues, it arose *in the course of* applying the latter, rather than *as the result* of applying the latter. What the example shows is that actors may manage to bring about the consequence they intend to bring about, but not in the way they thought that consequence would occur. The consequence is the result of a so-called wayward process[153] and cannot be described as something that is intentionally brought about. Indeed, in order to fit that description, an effect must, in addition to being intended, be brought about *in the right way*. Otherwise a wayward process is at play.

A wayward process is thus, aside from the more typical cases, another way by which the invisible hand can be operative[154]. On this suggestion, agents led by the invisible hand may be blind to the process by which an outcome is brought about, rather than to its occurrence. If this is so, agents involved in an invisible hand process may be moved by a recognition of the consequence of their action provided this consequence does not result from this recognition 'in the right way'. The definition of the invisible hand should be amended accordingly:

RA_2: an IHE is intended to explain a consequence of many actions performed by agents who may intend to bring it about as long as the consequence arises in the wrong way.

---

correspondence between the two is, however, completely unanticipated. The actors manage to bring about the consequence they intended to bring about, but not in the way they intended to bring it about. This is because the consequence is the result of a wayward process.

[153] Nozick clearly has this type of cases in mind when he says that "a theory would be interesting if it showed that, although everyone *was* aiming at a pattern, either their actions animated by that aim were not what produced the pattern, *or, if they did, that the pattern did not arise by the route everyone imagined — it was a side effect of their envisioned plans.*" (Nozick, 1994, note 7 [emphasis mine]).

[154] On Elster's classification, a wayward causal mechanism is not reducible to an invisible-hand explanation (nor, for that matter, to a design-intentional explanation). If, as I suggest, agents led by an invisible hand may either be blind to the effect of their action or to the way this effect occurs, a wayward causal chain turns out to be a way by means of which the invisible hand can operate.

4. Lewis conventions

Suppose now that the consequence does arise in the *right* way. Would it exclude the operation of the invisible hand? At first sight, it seems so. Agents led by the invisible hand are supposed to lack a clear view of what they do. A certain gap must separate what they are aware of and what they are actually doing. Since we previously let them be aware of what their actions add up to, we shall at least not allow them to be aware of the process by which it actually occurs.

Consider, however, the rule of driving on the right side of the road that prevails in Switzerland. There are many ways of explaining it. One is to explain it as the result of a policy. A legislator decides which side will be the driving side and punish those who violate that rule. Alternatively, the rule can be described as the result of a collective agreement. The drivers can decide which side they will drive on, unaided in this task by any third-party coordinator. Either way, the explanation relies on the ability of some agents to devise the rule.

There is however an alternative explanation, one that does not rely on a legislator, nor on some collective agreement. As I will show now, the theory that Lewis offers of conventions[155] is of that kind.

Conventions are, on Lewis' view, solutions to coordination problems. Deciding which side of the road should be the driving side, choosing something (e.g. cowry shells or coconut) as a medium of exchange, deciding who should call back when a phone conversation has been interrupted[156], limiting in some way the use of gas as weapon[157] are coordination problems. Two features characterize them. It is the sort of problem or dilemma which you have to deal with when (i) you have an interest in doing whatever the other does[158], whereas (ii) more than one option (more than one "coordination equilibrium", as economists would have it) could be a solution. There is no outcome by which agent's interests would be better served than another. Driving on the right side is not, *per se,* a superior option to that of driving on the left side[159]. What matters is that you and I end up choosing the same side.

A coordination problem is resolved when agents succeed in converging on one of the equally valuable options. According to Lewis, the capacity to form the right sort of mutual expectations plays, in this respect, a crucial role. In order to have their choice meeting on the same option, agents need to form some mutual expectations about each other's actions[160]. In Switzerland, agents expect each other to choose the right side of the road.

---

[155] Lewis, 1969.

[156] These three examples are offered by Lewis (1969, p. 9).

[157] Schelling explores this example in his *Strategy of Conflicts* (1960, p. 75).

[158] Unlike other social dilemma such as the prisoner's dilemma, agents facing a coordination problem do not have conflicting interests.

[159] Although these two options are commonly considered as equally attractive, driving on the right side actually is an inferior option because people tend to swerve to the left in an accident. Before the French revolution the best system prevailed and now only survives in the British Empire. The dilemma involved in choosing between left and right should therefore be construed as a Hi-Lo game rather than as pure coordination dilemma.

[160] As Lewis explains, mutual expectation splits into first-order expectations, as when I expect you to drive on the right side and *vice versa* and higher-level expectations about what you expect me to expect

Be that as it may, the left side would solve the coordination problem as well[161]. So the crucial question is: what makes the option of driving on the *right side* of the road the option of concordant mutual expectations?[162] Lewis argues that mutual expectations converge on the right side of the road because it is a more salient option than the left side. A salient option is an option that commands attention, one that "stands out from the rest by its uniqueness in some conspicuous respect"[163]. It is, in Schelling's words[164], a "focal point" and can be used as such as a coordination cue.

What makes an option salient? For an option to be salient it must be salient for many, and believed to be salient among those for whom it is so. Salience is a property that depends on infinite iterative mutual expectations. As Lewis says, "agents might expect each other to expect each other to expect each other to have [the] tendency [to pick the salient option] and act accordingly"[165]. In the same vein, he says that salient option is "unique in some way the subjects will notice, expect the other to notice, and so on"[166].

Lewis mentions precedent as a prominent source of conspicuousness. An option is eye-catching because we reached it last time. In the case of the driving convention, salience is a matter of precedence. Driving on the right side is what drivers are used to doing. It is a regularity that can rightly be expected to continue into the future. It is also a matter of being the side of the road on which it is legal to drive. Other factors can contribute to the salience of an option. What commands attention may, according to Schelling, "depend[s] on analogy, precedent, accidental arrangement, symmetry aesthetic or geometric configuration, casuistic reasoning, and who the parties are and what they know about each other."[167]

---

you to do, about what you expect me to expect you to expect me to do, and so on. Schelling also recognises such iterative interdependence in the formation of mutual expectations. "The best choice for either", he argues, "depends on what he expects the other to do, knowing that the other is similarly guided, so that each is aware that each must try to guess what the second guesses the first will guess the second to guess and so on, in the familiar spiral of reciprocal expectations" (Schelling, 1960, p. 87). Lewis acknowledges a practical limit the length of that spiral, however, namely the fourth level.

[161] The caveat to this claim is set forth in the previous footnote.

[162] Rational choice theory is in this respect markedly unhelpful. Take its central assumption, namely, the maximizing utility principle. Obviously, the propensity to choose the option that delivers the best payoff does not point to any side of the road. Indeed, any of the two combinations of options — left/left or right/right — allow us to maximize our expected utility. Moreover, rational choice forbids agents to be influenced by the way the options are labeled. This is because, on rational choice theory, what truly distinguish one option from another one is its payoff and not its description. "Driving on the left" and "driving on the right" could be respectively referred to as "%&*" and "§+/" without altering their respective worth. To be distracted by the apparent extra value that a label such as "driving on the right side" gives to one of the option it designates is to be the victim of an irrational "framing effect". It is to have one's attention focused on contingent matters, i.e. how an option happens to be labeled, rather than on what is essential to it, namely its payoff. Clearly, not only is rational choice theory unhelpful at explaining how agents solve coordination problems, it actually forbids us to use devices, i.e. the distinctive connotations that are attached to their description, that seem intuitively useful.

[163] Lewis, 1969, p. 35.

[164] "Most situations", Schelling claims, "provide some clue for coordinating behavior, some focal point for each person's expectation of what the other expects him to expect to be expected to do. Finding the key, or rather finding *a* key — any key that is mutually recognized as the key becomes *the* key" (Schelling, 1960, p. 57).

[165] Lewis, 1969, pp. 33-34.

[166] Lewis, 1969, p. 35.

[167] Schelling, 1960, p. 57. Take the prohibition on the use of gas as a weapon that belligerents of WWII sought to implement. "No gas" turned to be the option on which they tacitly managed to converge. What makes it salient? As Schelling explains, it is a choice that stands at one extreme of a

Gathering the various elements presented in the last section, we can say that, according to Lewis, a convention is a regularity of behaviour that is a solution to a coordination dilemma to which agents choose to conform because of its salience. In Lewis' words[168]:

> A regularity R in the behaviour of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that, in any instance of S among members of P,
>
> (1) everyone conforms to R;
>
> (2) everyone expects everyone else to conform to R;
>
> (3) everyone prefers to conform to R on condition that the others do, since S is a coordination problem and uniform conformity to R is a coordination equilibrium in S.[169]

5. An invisible-hand explanation

Lewis explicitly intended his theory to replace an explanation of conventions that assigns a crucial role to either a legislator or to a collective agreement. He says that legislators (or, in the present case, the Swiss highway patrol) play a superfluous role in the existence of conventions. To be sure, the prospect of receiving a fine for driving on the left side gives me an incentive to drive on the right side. The punishments operate as an external incentive that outweighs all reasons I may have to act otherwise. Driving conventions happen to be the result of policy enforced by sanctions that leave no room, it appears, for the mutual expectations identified by Lewis. Yet, as deterring as that fine may be, Lewis argues, what ultimately dictates my choice is what I expect others will do. In fact, our mutual expectations are more important than fines. It is especially true of driving conventions where staying alive is at stake. He says that "the punishments are superfluous if they agree with our convention, are outweighed if they go against it, are not decisive either way, and hence do not make it any less conventional to drive on the right"[170].

Lewis also aims to provide an explanation that dispenses with collective agreement. To be sure, all conventions could in principle originate in, and maintain themselves by, a binding agreement. They however need not be so. Lewis offers five different reasons for rejecting what might be called the agreement conception of convention.

First, agreements are redundant when agents have no interest in acting differently from the majority. Typically an agreement is "an exchange of formal or tacit promises". Promises are mainly called for, Lewis notes, when the action one promises to perform is both collectively beneficial and against the individual interest of the agents. It makes sense, for example, to make the promise to use less water

---

spectrum of possibilities that includes many less radical and possibly more convenient choices. Yet "no gas" gets its distinctive appeal by standing outside the continuous gradation in the various possible ways of limiting gas use. It became a focal point for both sides' expectations in the absence of any "natural" break between the many ways of limiting gas use.

[168] The present definition is an un-amended version to which Lewis subsequently brings various refinements, which are however not worth presenting for the present discussion.

[169] Lewis, 1969, p. 58.

[170] Lewis, 1969, pp. 45, 48.

during a drought because the un-conditioned inclination everyone has not to decrease one's consumption. In case like this, agents need to bind themselves to prevent themselves from the temptation to free ride, and promising is one efficient way to do so. But unlike prisoner dilemma-type situations, agents' interests are, in coordination situation well-served when they simply act on their mutual expectations. Promises therefore are superfluous and "an exchange of declarations of present intention" suffices[171].

Second, agreements are not strictly speaking solutions to a coordination dilemma. Rather they are ways to escape that dilemma. When I agree to drive on the right side, I make the promise to drive on it unconditionally. As a consequence, one option clearly becomes the promisor's dominant option, i.e. the option he promises to choose unconditionally, thus imposing it on all. The situation is no longer a coordination problem because the mutual expectations that define the latter are replaced by a one-sided expectation. Surely the coordination dilemma is eluded this way. But the ensuing regularity of behaviour is not a convention, since the latter is a *solution* to a coordination problem, rather than a way of evading it.

Lewis gets his inspiration for the third argument from Hume. Initial agreements, he says, may have taken place a long time ago. Yet an agreement made in the past ceases at one point to have any binding effect. It is too remote or does not have any direct effect on those who are not party to the agreement. At most, it motivates indirectly, that is, by inducing mutual expectations among those who inherited conventions based on their forefather's agreements. But mutual expectations are then what directly motivate us. If this is so, "a convention created by agreement is no longer different from one created otherwise: it bears no trace of its origin."[172]

Finally, the possibility that conventions originate in agreements is inconsistent with the Lewis account. Lewis defines agreements as promises to act *unconditionally*[173]. Conventions are however not based on intentions to follow them *no matter what*. On the contrary: agents follow them on the condition that others follow them as well. Agreements therefore cancel the condition of mutual expectations that characterizes Lewis' conventions. While an agreement may sow the seeds of a new convention, it only fully comes into bloom once the agreement ceases to affect our choice. As Lewis says, "we have a convention only after the force of our promises has faded to the point where it is both true and common knowledge that each would conform to some alternative regularity R' instead of R if the others did"[174].

It should now be clear that Lewis argues against two rival conceptions of conventions that both attribute a key role to a designing mind. Lewis does not show that these conceptions are empirically inaccurate. Rather he points to their conceptual superfluousness. Lewis does not argue that it is conceptually inaccurate to attribute a role to agreements. External punishments and agreements might very well play a causal role in the actual emergence and maintenance of conventions. What

[171] Lewis, 1969, p. 34.
[172] Lewis, 1969, p. 84.
[173] As Lewis observes (1969, p. 84), however, we may promise to conform to R *on condition that others do as well*. Most conceptions that derive the existence of conventions from an agreement do not however conceive the promises to conform to R to have such conditional character.
[174] Lewis, 1969, p. 84. This last argument, it must be noted, does not show the superfluity of agreements in regard to the existence of conventions, only its inconsistency with Lewis' conception of convention. Agreements are superfluous component of conventions only to the extent that, as Lewis claims, mutual expectations are at their core.

Lewis distinctively shows is that they however are inessential to their existence. He shows that if either a sanction or an agreement are in fact involved in a given case, they are uninteresting contingent facts about that convention. Because it satisfies the "no designer condition" of all invisible-hand explanations, Lewis' account can be taken as a good illustration of the latter[175].

Yet the driving convention is not, on Lewis' theory, an unintended consequence of actions. On his account, conventions are *solutions* to coordination dilemmas and the intentional dimension of a solution must not be forgotten. Something is a solution if at least one person is initially aware of the related problem, is willing to solve it, and sees it as a solution. Solutions are thus intentionally found. To say that driving on the right side is a solution to a coordination dilemma excludes the possibility that agents *un*intentionally converge on the option of driving on the right side of the road. Choosing the same side of the road is the agents' goal rather than a by-product of some other concern. On the Lewis view, the situation of agents when they follow a convention is certainly not a situation in which they find themselves by accident. Rather it is a situation they deliberately bring about. Using the framework formerly invoked, we can say that the description "driving on the right side of the road" picks up a rule to which agents intentionally conform. That this conformity is, first, mutually expected (and expected to be expected in the now familiar infinitely reiterated fashion), and that it is, second, conditioned on everyone's conformity only adds to its being an intended result. A Lewis convention is not something which agents "stumble upon", to use Ferguson's oft-quoted expression. It rather results from everyone's successful, individual intentions to create it.

A Lewis convention is thus not an unintended consequence, and therefore lacks what is largely recognized as an essential component of an invisible-hand consequence. What is more, no wayward causal chain is involved in the establishment of a Lewis convention. The latter is a consequence of many actions performed by agents (a) who have its establishment in mind, (b) who let its establishment enter into their reasons for acting, and (c) whose intentions is the non-deviant cause of that consequence. Our last revised account of the theory of the invisible hand, RA_2, turns out to be as unable as the standard definition to accommodate a Lewis convention.

The situation is, at this point, rather worrisome. Indeed, the distinguishing feature of invisible-hand explanations is now unfortunately invisible. Because they are allowed to be the non-deviant cause of the outcome to be explained, agents led by the invisible hand are hardly distinguishable from those who act out of collective agreement. But we assumed invisible-hand explanations to be different from social pact theories and to exclude such collective agreement. A way of drawing a line between the invisible hand and social pact explanatory theories needs to be found. The solution, I will show in the next section, is to distinguish between two modes of intention, namely an individualistic and a collective mode.

[175] Pettit (1993) and Aydinonat (2008) also consider Lewis' theory of convention as an invisible-hand explanation.

6. Acting jointly

It is at this point that Searle's example of the graduate businessmen case presented above can be helpful once again[176]. The graduate students, remember, are not acting together, although it is common knowledge among them that they are pursuing the same goal by the same means, that is, by acting selfishly. Afterwards, Searle offers a suggestion as to what they would have to do to act jointly. Let them "all get together on graduation day", Searle suggests in this regard, "and form a pact to the effect that they will all go out together and help humanity by way of each pursuing his own selfish interests"[177] and a genuine case of collective intentionality will be obtained. On Searle's view, social pact theories involve agents who are acting together — who are forming a common idea of how the reality should be shaped. Having this shared goal in mind, participants to the pact bind themselves by mutually agreeing to act so that their shared idea of social reality can be brought into effect.

Conventions can be approached according to these two perspectives. A convention (or, for that matter, any social pattern) can, on the one hand, be the result of many actions performed by agents who are acting individually, yet interdependently, and with the same purpose in mind. On the other hand, a convention can be the result of many choices undertaken by agents who are also acting interdependently and with the same purpose but who distinctively view themselves as acting jointly.

As will be argued later, a Lewis convention is of the first kind. It involves agents who, just like the graduate students in the initial situation, manage to establish the same rule, e.g. driving on the right side, without communicating and forming a pact to this effect. Remember indeed that Lewis explicitly intended his theory of conventions to dispense with the mutual unconditional promises which constitute an agreement. Agents need not form a pact, according to Lewis, and mutually bind themselves by means of a promise in order to give themselves a reason to conform to a convention. Moreover, just like the goal of helping humanity, which belongs to the graduate students' sphere of intendedness, driving on the right side is undoubtedly an intended result. Just like helping humanity, moreover, driving on the right side is not the result of a *collective* intention. Let the drivers meet and discuss the matter, however, and driving on the right side would be the intended consequence of a joint intention. If it makes sense to speak of a collective sphere of intention, driving on the right side would enter it.

I propose to draw the line that separates invisible-hand explanations from intentional-design explanations by using the difference between these two cases. A social pattern is the result of an invisible-hand process just in case it is the result of a process during which agents reason and act individually. Let the latter reason and act jointly, however, and the same social outcome would fall within the scope of an intentional-design explanation. RA_2 should be revised accordingly:

RA_3: an IHE is intended to explain a consequence of many actions performed by agents who may act with a view to producing the outcome to be explained, as long as they are acting individually rather than jointly.

[176] Searle, 1990.
[177] Searle, 1990, p. 407.

I believe that RA_3 captures what most invisible-hand explainers have — or should have — in mind when they reject the need for a "collective agreement" or for a "social pact". I find such a definition of the invisible hand superior to the standard account, that is, to an account that delineates its scope in terms of unintended consequences. It saves us from the task of resolving the various puzzles that surround the identification of unintended consequences. It also accommodates a relevant example of invisible-hand explanation, namely Lewis account of conventions, which would otherwise have been ruled out. It finally separates the invisible-hand theory of social reality from the social pact theory, by assigning a different kind of agency to each of them.

7. Salience reasoning reconsidered

Yet this new definition of the invisible hand can only be agreed upon if one accepts that a Lewis convention involves agents who do *not* act jointly. Only if the discovery of focal points is, on Lewis and Schelling's theory, not related to a we-intention will RA_3 be a compelling way of drawing the line between invisible-hand explanations and intentional-design explanations.

In favour of such an interpretation, one can argue that Lewis explicitly uses game theory as a framework for his theory of convention and the question "what should *we* choose?" cannot be formulated within the approach of game theory. The reason is that game theory postulates that no one other than individuals can choose. Individuals, it is claimed, are the only unit of agency conceivable so that speaking about a group mind can only be, at best, a *façon de parler*. Hayek, to take one of the most forceful advocates of this view, strongly argues against those who treat as natural objects collectives that are, on his approach, nothing more than mental constructions[178]. On his view, the idea of a social mind mistakenly attributes a personality to groups, revealing an inappropriate use of anthropomorphic concepts. To assume therefore that a group exists in its own right — that a group can reason and take decision — is to be guilty of a category error. It is to attribute to an abstract construction some capacities that only flesh and blood individuals can possess.

Still, it is not certain, or so some will reply, that Lewis' theory of conventions really excludes any we-ness. The idea of an irreducible plural subject is not the only way of construing collective intentions. There are other ways, compatible with Hayek's methodological individualism, of conceiving the we-ness dimension of plural agency. One way is to assign it to the content of the intention, rather than to its bearer[179]. Taking the group's perspective, it is argued, affects the *content* of the preference that is to be fulfilled as well as the kind of action by which it is expected to be satisfied.

[178] Hayek, 1942, 1973.

[179] There is a third approach according to which what makes an action a joint one is a matter of *modality*. On this view, there are two modes of things, namely a solo mode and a together-with-another mode. Searle seems to endorse this approach when he says that the reference to a group appears in each one of us under the primitive form: "we intend to…", rejecting the content-based approach of the form: "I intend and I believe that you believe that…", on the one hand, and an implausible world spirit floating above individuals, on the other. To my knowledge, Mathiessen (2002) provides the most comprehensive account of such hypothesis.

I believe that Postema endorses this conception of collective intentionality when he claims that a Lewis convention is based on an understanding of salience reasoning as a type of "common reasoning"[180]. Postema argues that discovering focal points requires the capacity to set aside what *one*, for one's own part, finds most striking (e.g. as an English driver visiting Switzerland, for example, driving on the left side is more salient to *me*) in favour of what *we* find salient (e.g. the option of driving on the right side will sound more conspicuous for some relevant *we* comprising me and the other Swiss drivers). Yet the argument explains why taking into account the particular cultural background of the agents with which one interacts is often helpful in finding the solution to a coordination dilemma. What the argument does not show however is why doing so exemplifies collective reasoning rather than individualistic reasoning.

There are admittedly passages where Schelling and Lewis present their theory of conventions as if it involved some sort of capacity to act together on the part of agents. Schelling says, for example, that salience reasoning requires a "meeting of minds" so that "some kind of collaborative or mutual accommodation"[181] can be achieved. Lewis says that the salient option will be found "by putting ourselves in the other fellow's shoes"[182]. These quotes suggest that agents cannot find the salient option unless they manage to attune themselves to each other's minds in a way that seems to be typical of joint actions.

Still it is as easy to find other passages where the individualistic tone patently prevails. Finding the focal points, according to Schelling, requires that agents ask themselves "What would *I* do if I were *she* wondering what *she* would do if she were *I* wondering what *I* would do if *I* were she...?"[183]. Agents represent to themselves the choice they should make in the first person singular, rather than in the first person plural, conveying the idea that they are not really acting *together*. Surely, agents must refer to the choice the others are likely to perform in order to make their own. But discovering the salient option remains an individual task, in spite of the reference they make to others. In the same vein, Lewis depicts the process by which a convention arises as one "in which *one* person works out the consequences of his beliefs about the world — a world he believes to include other people who are working out the consequences of their beliefs, including their belief in other people who…"[184]. Lewis also describes the agents involved in this kind of reasoning as

[180] Postema, 2008.
[181] Schelling, 1960, p. 83.
[182] Lewis, 1969, p. 27.
[183] Schelling, 1960 [my emphasis].
[184] Lewis, 1969, p. 32 [emphasis in original]. Some notice that this infinite regress of interdependent expectations is a trap into which anyone trying to coordinate on the basis of the salient option falls (cf. Gilbert, 1989, Sugden, 2000, Tuomela, 2002, p. 390, Schmid, 2009, chapter 6). The problem is the following one. The reason agents have to act on the salient option is always conditional on what they expect their partners will do. I have a reason to play my own part in the combination of the salient option — to drive on the right side — *if* I have reason to expect you will play yours. But you are in the same position, having a reason to drive on the right side *if* you have a reason to expect I will do so as well. The fact that a good outcome — no collision happens — would be reached *if* both did something cannot by itself be a reason for either one individually to do it. The argument begins with a conditional about what will happen if each player acts on the salient option, but there does not seem to be any rational way of getting rid of this conditional. This leads to an infinite regress of the type "I expect my co-player to expect me to expect..." that provides neither player with any rational justification for choosing the right side rather than the left. Like Buridan's ass, one ends up picking none. In sum, the fact that a particular combination of strategies is a focal point gives neither agent a reason to play his part in it. Within the logic of individual rationality, salience reasoning is impotent.

"*windowless monads* doing [their] best to mirror each other, mirror each other mirroring each other, and so on"[185].

Yet a Lewis convention, it can now be replied, involves agents who are acting *interdependently* — their willingness to do their part depends on the other's willingness to do his — and the latter is often considered as a mark of joint action[186]. As a counter objection, one can, following Matthiessen, note that interdependency is not, first of all, a sufficient condition of joint actions since it may be a feature of individual actions. To illustrate this point, Matthiessen puts forward the case of Susan and Bill the spiteful theatregoers. Bill hates Susan and Susan knows that her appearance at the play will ruin Bill's enjoyment. Similarly, Susan hates Bill and Bill knows that his appearance at the play will ruin her enjoyment. So, the following conditions have been met (a) Susan intends that Bill and she go to the play, (b) Bill intends that Susan and he go to the play, (c) Susan intends to go because Bill is going and Bill intends to go because Sue [sic] is going, (d) there is common knowledge between Bill and Susan about (a), (b) and (c). Susan and Bill may very well end up going to *Death of Salesman*. There would be something odd in either of them saying that "we" intend to go. As this example shows, agents may have interdependent intentions, they may pursue the same end, and this may be common knowledge among them, while they do not share a collective intention[187].

Nor is interdependency, as Mathiessen further notes, a necessary condition of joint action. The case she offers in defence of this point is one where I intend to go see *Death of a Salesmen* with you, but still intend to go even if you cancel. Thus, my performing my part of this collective action is not dependent on my belief that you will perform yours. As Mathiessen explains, there may be cases where you and I share a collective intention to do something, even if I would still do it on my own without you.

The idea that agents who are involved in forming a Lewis convention act individually—even if they act strategically, interdependently, and with the same goal in mind—should now, I hope, sound plausible. This is not to say that conventions could never rest on common reasoning. Whereas Lewis conventions are not of that kind, conventions could alternatively be construed as the result of agents acting with the shared purpose of bringing them about. Reflecting on what agents would have to do in order to be involved in such a process is another way of delineating the boundary condition of the invisible hand.

Sugden's idea of team-directed reasoning[188] is in this regard helpful. Just like Lewis, Sugden views conventions (or social rules) as solutions to coordination dilemma and also uses the game theoretical approach. Unlike Lewis theory, however, Sugden allows groups, or, as he says, "teams", to be legitimate units of agency. On his view, the choice between driving on the right or on the left side should be framed as a problem which teams, rather than individuals, must tackle. The idea is that if I

Salience reasoning is explanatorily worthless, it has further been argued, insofar as it is conceived as *individualistic* reasoning. As long as each agent is only allowed to represent what she should do in the perspective of the first person, she will remain impotently trapped in the vicious circle of "ifs". Let them ask themselves "what shall *we* do?", however, and they will find an easy way to avoid this regress. (Cf. Tuomela, 2002, pp. 395-396, Sugden, 1993, 2000, 2003, Gold and Sugden, 2007, and Schmid, 2009).

[185] Lewis, 1969, p. 32 [emphasis original].

[186] Miller, 1995.

[187] Note that this argument is also against putting we-ness into the content of agents intention.

[188] Sugden, 1993, 2000, 2003, 2007.

consider myself as a member of a team[189] that includes you as the other member, I will not conceive of the decision problem as a problem *for me* but as a problem *for the team*. Team-directed reasoning involves two tasks. One is to identify the team's most highly ranked preference and the other is to determine what individual actions will best promote the realization of that goal. As Sugden says, these two tasks are analogous to what is going on when one takes an individualistic perspective. As he concedes, "the role that team preferences play in my theory of team agency is essentially the same as the role that individual preferences play in the standard theory of individual agency" (2000, 176). Be that as it may, taking the team's perspective must *somehow* be different from taking the individualistic perspective. "Team agency", as Sugden conceives it "is not reducible to individual agency as that is represented in rational choice theory"[190]. He argues in favour of a reading of team preference "in which the preferences of a team are not necessarily reducible to, or capable of being constructed out of, the preferences that govern the choices that the members of the team make as individuals".

Taking the perspective of the team implies that, first, I will be inclined to identify what the team's goal is and that I will, secondly, have to figure out what action on my part maximizes the likelihood of reaching that goal. Why is team-directed reasoning a *collective* way of deliberating? Sugden gives the following reasons. First, when agents are engaged in a team-directed reasoning activity, they consider their individual action in terms of "playing their part" in the combination of actions that is best for the team[191]. By contrast, when someone acts individually, he does not conceive of what he does as playing his part in a combination of actions[192]. Rather he conceives of his choice as a means of satisfying his own preference[193].

There is a second feature of team-reasoning that makes it distinctively non-individualistic: it generates recommendations for action that are not conditional on the actor's beliefs about what the other individuals will do. In this respect, team-directed reasoning is quite different from the strategic reasoning that conventional game theory serves to model. Interdependency, i.e. the rule "I will if you will" that partly defines a Lewisian convention, thus seems not to be a feature of the way conventions are jointly discovered and followed[194].

---

[189] Choosing the image of a team to reflect on joint action might sound like an awkward choice. Teams are typically coached so that agents acting as team members are not acting autonomously. They merely follow the coach's recommendation. Joint actions are however generally considered as being performed by agents acting under no one's command but their own. Joint actions cannot involve agents who, as team members, are monitored by someone acting externally. Against this view, Schmid argues that it confuses autarky, i.e. the idea that all individuals participating in a joint action are agents in their own right, with autonomy, i.e. the idea that all behaviour has to bottom out in that individual's own volitions (Schmid, 2008, p. xv).

[190] Sugden, 2003, p. 176.

[191] Sugden, 2008, p. 184; Sugden, 2000, p. 187.

[192] Note, however, that on such criteria, Lewis describes agents acting jointly. Indeed, he frequently describes agents' choices as ones by which each play their part in the equilibrium.

[193] "In the standard theory, the individual appraises alternative options in relation to…her preferences, given her beliefs about the actions that other individuals will choose. An individual engaged in team-directed reasoning appraises alternative *arrays of action by members of the team* in relation to [the] team-directed preferences (Sugden, 2000, 187). Consider also this quote: "When an individual reasons as member of a team, he considers which *combination* of actions by members of the team would best promote the team's objective, and then performs his part of that combination." (Sugden, 2003, p. 167).

[194] The lack of conditionality in the conclusions of team-directed reasoning also allows an escape from the trap of the infinite regress (Sugden, 2000, p. 191). However, in a recent development of his theory

The activity of team-reasoning could very well represent the limit that agents led by the invisible hand shall never cross. It is where the line dividing invisible-hand accounts from rival accounts is to be found and RA_3 should be redefined accordingly:

RA_4: an IHE is intended to explain a consequence of many actions performed by agents who may act with a view to producing the outcome to be explained, as long as they represent to themselves this intention in a solo mode rather than as a together-with-the-other mode.

Conclusion

In this chapter, I have argued against the conventional way of defining the scope of the invisible hand in terms of unintended consequences. I have shown that precluding the existence of designers, be it an external ruler or a group of agents acting jointly to produce some outcome, is not sufficient to ensuring the unintendedness of the outcome to be explained.

On the revised account that I propose, agents led by the invisible hand should not be prohibited from having the ultimate outcome of their actions in mind. Nor should they be prevented from having the production of that outcome as part of their intention. They should only be prevented from having the production of that outcome as part of their collective intention. To have a role in an invisible-hand process, agents may act with a view to contributing to the occurrence of the pattern to be explained, provided they see what they do as an aggregation of their individual actions rather than as something they jointly perform. They should, in other words, be allowed to have the intention to do their part in bringing about the outcome to be explained, inasmuch as they see "bringing about the outcome to be explained" as an effect they intend to produce *individually*. Let agents start thinking of themselves as members of a collective agency, however, and no invisible-hand explanation will be available.

(Gold & Sugden, 2007), Sugden seems to have changed his view on this matter. A team reasoner who identifies with a group, he says, stands ready to do her part in a joint action. "But she does not necessarily take this objective as hers in the stronger sense of wanting it even if other members of the group do not reciprocate" (Gold and Sugden, 2007, p. 132).

III. When The Bell Rings

Introduction

Invisible-hand explanations are commonly seen as providing a conjectural account. They describe how some social phenomena could take place rather than how they actually take place. A conjecture appears however to be different from an explanation. An explanation is generally expected to expose the factors that are at work in the bringing about of some patterns. It is supposed to inform us about the *actual* causal story of that pattern. A conjectural account, by contrast, uncovers a *hypothetical* causal story. Because hypotheticals remain agnostic about the truth of the antecedent, how they can possibly provide any explanatory information about the real world is therefore a mystery. Hence the widespread suspicion that invisible-hand explanations stir up with regard to their explanatory power.

Specifying their conjectural character, it is argued, will alleviate the suspicion. On this suggestion, an invisible-hand explanation does not merely describe how some social phenomena could take place[195]. It shows how it would have taken place, had it not taken place the way it actually did. This is, for example, how Nozick conceives of the manner in which an invisible-hand explanation accounts for the reality that it describes. It is a "potential explanation", he claims, which he characterizes as an explanation that "would be the correct explanation if everything mentioned in it were true and operated"[196]. Nozick further distinguishes between three variants. A potential explanation can either be law-defective, fact-defective and/or process-defective. The last of these he defines in the following way: "A potential explanation that explains a phenomenon as the result of a process P will be defective … if some process Q other than P produced the phenomenon, though P was capable of doing it. Had this other process Q not produced it, then P would have."[197] An invisible-hand explanation, in other words, describes the pre-empted cause of phenomena. It points at the cause that would have been at work had some other cause not been operative instead.

The crucial question is: what then, if anything, is explained this way? What sort of explanatory information does the uncovering of the pre-empted cause of a given social pattern convey? Here is the answer. Showing that there is a series of causes that are ready to operate, in case the actual cause of the social phenomenon ceased operating explains its resilience, that is, why a given social pattern maintains itself in many circumstances. Ullmann-Margalit recognises this special kind of explanatory power that invisible-hand explanations have. According to her,

> even if the invisible-hand explanation turns out not to be the correct account of how the thing *emerged*, it may still not be devoid of validity with regard to the question of how (and why) it is *maintained*. … The availability … of a cogent invisible-hand story of how the pattern in question could have arisen — given the specific circumstances, some common-sense assumptions concerning the drives of the individuals concerned, and the normal course of

[195] I may here incorrectly suggest that there are only two possibilities — the actual one and the one that is described in the invisible-hand explanation. But the latter, as it will become clear, refers to many possibilities inasmuch as it assumes a range of individual preferences falling somewhere on a spectrum between extreme selfhishness, on the one side, and moderate selfishness, on the other side.
[196] Nozick, 1974, p. 7.
[197] Nozick, 1974, p. 8.

> events — may, I believe, contribute to our understanding of the inherently self-reinforcing nature of this pattern and hence of its being successful and lasting.[198]

The most comprehensive development of this idea is however to be found in Pettit's idea of the virtual operation of self-interest. In brief, his proposition is to say that an invisible-hand account explains the resilience of a social pattern inasmuch as it shows how the latter would be obtained all the same, even if, having their self-interest threatened, agents ceased to act in the selfless way in which they routinely behave.

The model of virtual self-interest is a revised version of rational choice theory. The first section in this chapter presents the standard version of that theory and the problem it raises. The second section presents Pettit's revised account. A few paradigm invisible-hand explanations are reformulated according to Pettit's amended version of rational choice theory in the third section. The fourth section shows why Pettit's theory allows invisible-hand explanations to be potential explanations, rather than mere logically valid conjectural accounts. The last section shows that the model of virtual self-interest explains the resilience of social pattern only under very restrictive conditions. The main implication concerning the scope of Pettit's theory is brought to light in conclusion.

1. Rational choice theory: the conventional view

Rational choice theory comes in many different versions. There are however a few ingredients that are at the core of all of them and that make up the conventional view. In order to appreciate the nature, extent, and relevance of Pettit's amended conception, we will first set out what can be considered as the conventional one (1.1) and the problem it raises (1.2).

1.1. Two central assumptions

Pettit identifies two central assumptions within rational choice theory. The first, he calls a "process-centred assumption"[199]. The assumption is *process*-centred, I presume, since it construes action as the final stage of a process in the course of which agents deliberate about what to do. Beliefs alone, on this view, do not play any motivational role in this conception of agency, unless they are attended by a desire. The assumption has, in turn, many corollaries. One is that agents act on the basis of their *own* desires rather than according to other people's desires[200]. The assumption is thus process-*centred*, inasmuch as it assumes that all actions are based on its performer's desire. Another corollary specifies the way in which desires issue in actions. Agents are in particular recognised as having more than one desire and as being able to rank them according to the utility that realizing those desires would deliver. The utility of a desire is variously defined in terms of the pleasure, of the material reward, or in terms of any

[198] Ullmann-Margalit, 1978, p. 275.
[199] Pettit, 2007.
[200] Pettit, 2000, p. 234. While considering the idea as commonsensical, Pettit recognises various attempts to challenge it. Pettit mentions in this respect Sen (1982) to which Searle (2001) and Schmid (2005) could be added.

subjectively valued state that is brought about by the satisfaction of the desire. It is a central assumption of the conventional conception of rational choice theory that agents always perform the action that delivers the highest amount of utility.

A second set of "content-centred assumptions" specify the sorts of things that agents desire. On Pettit's view, rational choice theory assumes a certain level of egocentricity. Self-regarding desires are, in particular, assumed to be stronger than other-regarding ones. It is taken for granted that conflicts between what will satisfy my interest and what will satisfy yours always issue in favour of the former kind of desiderata. Agents are thus inclined to perform the most selfish action available[201].

Some will perhaps be surprised to find the content-centred assumptions at the core of the conventional view of rational choice theory. Among its advocates[202] and critics[203], indeed, many explicitly exclude this from its foundation. Only the process-centred principle of utility maximisation predicting that people do what they subjectively prefer to do is retained, leaving open the possibility of altruistic actions if the latter happens to be what agents prefer doing. To be sure, the utility language, inherited from the utilitarians[204], may suggest that rational choice theory also specifies *what* people want and that the painful cost associated with most altruistic actions will therefore be avoided. Yet most theorists hold that rational choice does not involve this restriction. Utility, they argue, shall be treated as an empty placeholder, allowing for any kind of mental or physical states, as along as they are theirs, to be the object of maximisation. If this is so, rational theory may very well accommodate people who rank the well-being of others above theirs. They are no less utility maximisers than those who have an exclusive concern for their own well-being.

Against this widespread view, Pettit regards the content-centred assumption as a genuine feature of rational choice theory. As it turns out, most applications of that theory support his interpretation. By and large, economists do assume that agents are thoroughgoingly selfish in their predictions and explanations, notwithstanding the popularity of experimental economics whose main project is to cast doubt on the explanatory power of such assumption[205]. More crucially, Pettit argues that economists are implicitly committed to the egoism assumption in their theoretical tools, despite the way they explicitly deny any such commitment. For example, it is an assumption without which many economic tools and notions, such as the downward-sloping demand curve and the Paretian assumption, would simply not be plausible[206]. The indifference curve, it can also be added, is based on the idea that having more goods for me is better than less. Pettit is right to say that the predominant framework for

[201] Pettit, 2001, p. 79.
[202] For example, cf. Hausman & McPherson (2005).
[203] For example, Elster (2007).
[204] In drawing a parallel between rational choice theory and utilitarianism, I may seem to be *wrongly* assuming that proponents of both theories presuppose that agents are inherently selfish. I do not make this assumption. Utilitarianism and rational choice theory have a complex relationship which can be in part elucidated in the following way. On the one hand, both theories can be considered as cousins inasmuch as they are (i) normative, i.e. they both purport to elucidate the way agents should act (although rational choice theory *also* aims at describing how they act), (ii) consequentialist, i.e. they both claim that all that matters is the consequences of one's actions (in contrast to, say, its intrinsic value or obligatoriness), and (iii) defending the truth of maximization, i.e. they both claim that more of good consequences is always better rather than less. On the other hand, the two theories diverge in regards to who the good consequences should benefit. Whereas the utilitarians conceive what is to be aimed at as the greatest good of the greatest number, the rational choice theorist takes the individual agent's good as the unit of maximization.
[205] Cf. for example Fehr (2002) and footnote 208 below.
[206] Pettit, 2001, p. 80.

rational choice theory commits its practitioners to the egoism as a content-related assumption. Whether they like it or not, economists are indeed committed to the assumption that people's self-regarding desires are generally stronger than their other-regarding ones.

1.2 The problem with the conventional view

A common way to argue against rational choice theory is to stress its inaccuracy. Although the process-centred assumption is not immune to criticism, Pettit only objects to the content-centred assumption, which, on his view, "flies in the face of common sense [and] conflicts with our ordinary assumptions about how we each feel and think in most situations"[207]. Indeed, a broad range of human interaction shows that human beings are not of predominantly self-regarding concern. Everyday experience shows that we are moved by considerations that are far from being exclusively egocentric. In particular, fair play, friendship, loyalty, kindness, politeness, honesty, and frankness are also motives operative in many human situations[208].

The discrepancy between common sense and rational choice theory raises some doubt about the ability of the latter to be of explanatory value as regards real world situations. Rational choice theory, Pettit says, is subject to the problem of the "empty black box":

> The mind postulated in rational-choice theory is that of a relatively self-regarding creature. But the mind that people display towards one another in most social settings, the mind that is articulated in common conceptions of how ordinary folk are moved, is saturated with concerns that dramatically transcend the boundaries of the self. So how can we invoke the workings of the economic mind to explain behaviour, when the "black box" at the origin of behaviour does not apparently contain an economic mind?"[209]

As an inaccurate description of how agents deliberate and act, how could rational choice theory be of any interest? The question has received many different answers. Before looking at Pettit's, let us review of few of them. In his discussion of the subject, Elster recognises three reasons to assume selfishness when offering an explanation. First, selfishness must be assumed because, unlike altruism, it could be a universal inclination. As Elster says, "the goal of the altruist is to provide others with an occasion for selfish pleasures … If nobody had first-order, selfish pleasures, nobody could have higher-order, altruistic motives either"[210]. In other words, while the altruists need the selfish to be who they are, the reverse is not true[211]. Secondly, taking

[207] Pettit, 2001, p. 75.

[208] Rational choice assumptions also are at odds with results from experimental economics. The ultimatum game (Guth *et all.,* 1982) shows that quite a few people are ready to forgo a substantial sum of money for the sole benefit of denying a larger sum to an anonymous stranger who has treated them ungenerously, even if they have no prospect of interacting with him again. Other evidence, drawn from laboratory experiments, shows that behaviours are determined to a substantial extent by motives other than self-regarding preferences. For example, the observed propensity to punish defectors in these public-goods experiments (Fehr & Gachter, 2000), even at cost to oneself, seems to be at odds with the selfish assumption of rational choice theory.

[209] Pettit, 2000, p. 238

[210] Elster, 1990, p. 45.

[211] It can be replied that a world exclusively composed of altruists is not impossible if its selfless inhabitants *wrongly* believe that a few among them can selfishly benefit from their magnificence. Moreover, Elster implicitly requires from the altruist that he does not take pleasure in the benevolence

selfishness as the default position spares one of the humiliation in being corrected by a more sophisticated colleague showing that the behaviour under scrutiny is perfectly compatible with the self-interest assumption[212]. Another reason is that we often do not know for sure whether a certain apparently disinterested behaviour really is based on a genuine disinterested motivation or not. Caution requires us to presuppose selfishness while being ready to revise the assumption in light of further contradictory evidence[213].

These are however admittedly weak reasons to take selfishness as granted when offering an explanation. As we shall now see now, Pettit's defence of selfishness as a virtual cause of behaviour is more compelling.

2. The virtual reality of *Homo Economicus*

2. 1 The virtual presence of self-interested thoughts

Pettit proposes a way to save rational choice theory from the charge of inaccuracy and his solution is rather simple. It is to conceive self-regard as a *virtual* cause of behaviour rather than as its *actual* cause[214]. Self-interest, according to Pettit, does not have to figure as what actually motivates agents to act the way they do. It does not have to be the actual cause of human behaviour because it might, as he says, "have a virtual presence in people's minds"[215].

As Pettit explains, having a virtual presence in the generation of an action is compatible with having no impact on this action. Self-interest may endlessly remain the virtual, pre-empted cause of some behaviour. In other words, a virtual cause is not the same thing as an actual mental state, an actual cause, the content of which is hypothetical. However, it is also possible that, on some occasions, the virtual cause ceases being so and takes the place of the actual cause of behaviour. It happens in particular when agents are moved to see their circumstances from the point of view of their self-interested preference. They will be given the motivation to reconsider their action when their personal advantage gets compromised. Loyalty, kindness, politeness, honesty or straight talk will then give way to mercenary thoughts. On Pettit's view, selfless preferences only have a bearing on what people do insofar as people's self-regarding interests remain protected. It is when the latter is jeopardized, however, that

towards his or her peers, at the risk of losing her altruistic character trait. The constraint is obviously too demanding.

[212] Only assuming the possibility of altruism reflected a sentimental naïveté would it appropriately trigger the mockery of the sophisticates.

[213] Elster, 2009, p. 30. It can here be replied that caution rather suggests further inquiries be made and requires one to assume nothing until new evidence presents itself.

[214] Note that this is not the only amendment that Pettit imposes on rational choice theory. He also criticises the theory for having a restricted view about what agents desire. On the conventional conception, desires are mainly conceived in terms of desires of economic advantages. Self-regarding desires, he argues, may extend to other goods. They may pertain to the good of being loved, of being well regarded and of being socially accepted (Pettit, 1990, Brennan and Pettit 1993). Pettit strongly defends the possibility of "attitude-dependent goods" as he calls them, that is, of goods that one enjoys in virtue of other people's attitudes toward us rather than in virtue of action performed by oneself. Attitude-dependent goods are in particular the goods we enjoy in virtue of the admiration, approbation, good opinion, or esteem of others. On the basis of this amendment, Pettit (together with Brennan, 1993) defends the possibility of a behavioural control system he calls the "intangible hand", which he sees as operative every time agents unintentionally sanction each other by forming a good or a bad opinion of each other's actions.

[215] Pettit, 2007, p. 279.

people start thinking in conformity with rational choice assumptions. They start calculating the personal losses and benefits that their behaviour results in.

There are two ambiguities in the way I have presented Pettit's idea of virtual self-interest that needs to be cleared up. One way to interpret it is to say that we weigh moral and cultural considerations *along with* and against self-interested considerations and, under certain circumstances, the self-interested considerations come to outweigh the others — those circumstances being circumstances where self-interest is not somehow maximised.

This is not however what Pettit seems to be suggesting. He rather seems to claim that moral and cultural considerations are *ignored* when the satisfaction of self-interest moves potentially below some threshold. So on this view there are threshold effects that alter the framework for our decision-making deliberations — a framework within which one now only measures considerations in terms of maximising self-interest. This threshold-triggering "reframing" point something quite distinct and more radical from the mere weight-related maximisation point referred to as in the previous paragraph.

Now the second ambiguity. According to Pettit, we stop caring about other regarding goods if and when we come back down to the threshold where self-interested goods start having marginal utility again (either because we got greedier, or we lost some goods, or face the prospect of losing some of them). It might be attractive to assume a symmetrical claim regarding the way other-regarding considerations bear on our decision. On the related construal, selfish goals eventually hit a marginal utility of zero (where they are fully satisfied), only after which level other-directed goals kick in. In other words, we act according to rational choice theory up until the point where self-interested goods stop having any marginal utility for us, leading us to start being interested in other-regarding goods. On my reading, however, Pettit's idea of virtual self-interest does not comprise such a symmetrical claim. Virtuality only pertains to selfishness in a way that makes being moved by other-directed motivations the only defeasible state of affairs.

It is finally worth stressing that the model of virtual self-interest involves the presence of signals that catch people's attention. Those signals tell us when our behaviour ceases to be to our advantage. For example, the fact that others fare better brings to mind the possibility that moral or cultural considerations should give way to more self-interested concerns. The model of virtual self-interest assumes, also, the readiness on the part of agents to listen to these signals. Human beings are known to be very likely to discover and exploit any opportunity to further their own interests even at the expense of others' interest. Both the presence of signals and the familiar willingness to pay attention to them are well-known facts that give credibility to the model of virtual self-interest.

Pettit's model of virtual self-interest can be summarized in the following way. Agents pay no actual attention to relatively self-regarding considerations, but those considerations have nonetheless a relevance to how they behave. They are virtually present in the sense that if their behaviour turns out not to be personally advantageous, they will start reasoning in self-regarding terms.

Pettit's version of rational choice theory introduces a significant twist to the conventional theory by making its central assumption a description of how agents

virtually behave rather than how they *actually* behave[216]. If self-regarding considerations have a virtual presence, they are not the causes of actual behaviour. Rather they are their "standby causes"[217], "potential causes" or, as Pettit also calls them, "causes of a would-be variety"[218]. Making self-interest a virtual concern rather than an actual concern will protect rational choice theory from the criticism presented in the foregoing. It is no longer relevant to stress the discrepancy between how agents really behave and how the theory falsely represents them. For the theory does not aim to describe the reality, but its virtual counterpart.

2.2 A subjectivist theory

Yet Pettit remains also faithful to many other claims that are at the core of the conventional version. In particular, rational choice theory remains, on his conception, a subjectivist theory. Nobody other than the agent herself in position to tell when the level of self-sacrifice that she endures while acting as a loyal, faithful and friendly agent, is too high to be tolerable. Pettit admits that the threshold at which agent's interest is compromised varies from individual to individual[219]. In other words, rational choice theory predicts that agents will try to maximise their selfish interests, while having nothing to say about what must count as the situation in which this self-interest is satisfied. It is a matter that the theory cautiously lets each of us assess subjectively[220].

People's views vary considerably as to when their self-interest is threatened. The virtual disposition to pay attention to one's self-interest becomes operative under widely different conditions from one agent to another. While some agents will sense that their personal interest is jeopardized as soon as the slightest loss occurs, others will wait for a large amount of self-sacrifice before reaching the same conclusion. These variations may even be encountered in one and the same agent, depending on her circumstances or moods. But these variations have no hold on Pettit's version of rational choice theory. The theory provides room both for the robust altruist and for her less kind-hearted counterpart. Indeed, the fact that they do not have the same view about when their self-interest is threatened does not prevent the latter from playing its virtual role. In sum, an inappropriately resented false threat may have the same altering effect on one's way of deliberating as a true one. Leaving the notion of self-interest as

[216] In other articles (1992, 1996), Pettit defends the class of functional explanations using the same twist. He first observes that most functional explanations in social science are based on an implicit but yet mysterious mechanism of an institutional selection. Yet functional explanations remain valuable, he argues, once the unarticulated mechanism is seen as being virtually at work.

[217] Pettit, 1996.

[218] Pettit, 2000, p. 242.

[219] Pettit, 2007, p. 271.

[220] To be sure, there is a substantive way of assessing when someone's self-interest is threatened irrespective of what agents personally believe about this matter. One may, for example, argue that agents' personal interests are threatened when their health, their longevity, their income, or any other objectively measurable good, declines. Alternatively, one may call for a Kantian reading of what is in all agents' personal interests to do in terms of autonomous actions. Whatever their force, however, such answers will not be proposed by Pettit nor by any rational choice theorists, who carefully avoid endorsing any objective conception about what agents should self-interestedly prefer. There are various reasons for remaining formal on this matter. One is that any substantial view about what is in agents' interests to prefer will sound too paternalistic. Letting agents autonomously decide what it is in their interest to do coheres with a liberal attitude which most rational choice proponents enthusiastically endorse. Moreover, a substantial view about what agents should prefer would also be controversial. Assuming nothing on this matter is a tactical way of accommodating everyone's various intuitions. There is, finally, the problem of explaining a given behaviour as the result of what is objectively in the interest of its author to do even though she is not aware of it.

an empty place holder allows all sorts of behaviour — even apparently extremely other-regarding one — to count as being virtually backed-up by self-interested considerations.

2.3 Resilience explained

The crucial question is: what information about the social pattern that is to be explained do we get from assuming virtual selfishness? What is the explanatory power of citing a standby cause? To answer this question, Pettit distinguishes between three types of *explananda*. We can first be interested in explaining the emergence (the appearance, occurrence, bringing about, etc.) of a social pattern. We may secondly be interested in explaining its continuation (its survival, reproduction, maintenance, preservation, duration. etc.) or we could finally be concerned to explain its resilience (its self-maintenance, robustness, stability, etc.)[221].

To explain the emergence and maintenance of a social pattern, one will have to trace its actual causal story. But an invisible-hand explanation makes no such attempt. Rather it shows "why it is such that for a variety of possible scenarios, possible ways the world may go, it remains a fixture"[222]. An invisible-hand explanation shows that a very large number of initial states will evolve in such a way that the system is likely to end up in the outcome state that we wish to explain. It explains why, in other words, the pattern remains in spite of possible disturbances.

From this new perspective, the explanatory capacity of an invisible-hand explanation seems both more limited and broader than from the conventional perspective. On the one hand, it seems more limited because, having the power to show only the resilience of a social pattern, it turns out to be useless in regard to its emergence or persistence. Its explanatory power seems, on the other hand, broader. Sober endorses this view in his discussion of what he calls "equilibrium explanations". As he shows, they offer no account of the actual causal sequence that led to the existence of a given social pattern. Such an account however "situates that actual trajectory (whatever it may have been) in a more encompassing structure"[223].

2.4 The ball rolls metaphor

Pettit uses a number of lively metaphors to present his version of rational choice theory. One of the most telling is the alarm bells:

> There are certain alarm bells that make [agents] take thought to their own interests. People may proceed under more or less automatic, cultural pilot in most cases but at any point where

[221] Pettit also mentions a practical reason to focus on the resilience of social phenomena specifically. Learning that an institution is resilient is a crucial piece of information for institutional designers. The durability of an institution or its resilience in the face of changing preferences and shifting commitments is indeed an advantage that they will try to add to the institutions that they design. Now such durability or resilience is likely to be obtained if they assume selfishness on the part of the agents. Assuming that agents are selfish is not to assume too much on the part of people. In particular it is not assuming too much, to say the least, concerning their moral aptitudes. Institutions that welcome knaves, as Hume argues, are institutions that pass the resilience test. Against this view, some have argued that assuming selfishness will have the counterproductive effect of crowding out agents who are not selfish (cf. Frey, 1997).

[222] Pettit, 2007, p. 83.

[223] Sober, 1983, p. 207.

> a decision is liable to cost them dearly in self-regarding terms, the alarm bells will tend to ring and prompt them to consider personal advantage; and heeding considerations of personal advantage will lead people, generally if not invariably, to act so as to secure that advantage: they are disposed to do the relatively more self-regarding thing.[224]

Far from playing only a rhetorical role, metaphors do have an argumentative import in Pettit's reasoning. For example, the rolling ball metaphor illuminates in a vivid way the explanatory power of standby causes:

> Imagine a little set-up in which a ball rolls along a straight line — this, say, under Newton's laws of motion — but where there are little posts on either side that are designed to protect it from the influence of various possible but nonactualized forces that might cause it to change course; they are able to damp incoming forces and if such forces still have an effect — or if the ball is subject to random drift — they are capable of restoring the ball to its original path. The posts on either side are virtual or standby causes of the balls' rolling on the straight line, not factors that have an actual effect.[225]

The posts are not responsible of the rolling down of the ball. Nor do they explain the continuation of its straight course. They however ensure that the ball

> sticks to more or less that straight line under the various possible contingencies where disturbance or drift appears. They explain the fact, in other words, that the ball's straight trajectory is not something fragile, not something vulnerable to every turn of the wind, but rather a resilient tendency: a tendency that is robust under various contingencies and that can be relied upon to persist.[226]

As a metaphor, the rolling ball setup is meant to cast light on the kind of explanatory power of the self-interest assumption. The analogy goes as follow: just as pinball posts ensure that the ball rolls straight no matter what, the selfish assumption insures that an agent's current behaviour is very unlikely to be modified. Section 5. will explain why the two situations might not be as strictly analogous as they seem.

3. A major style of explanation in social sciences

Pettit's revised conception of rational choice theory will be even more compelling if, in addition to being internally consistent, it also fits the practice of social scientists. We should therefore be able to reformulate all conventional rational choice accounts as providing a virtual account of social reality.

3.1 The slavery system: why it persisted

[224] Pettit, 2000, p. 240. As Otto Bruun pointed out to me, the ringing bells metaphor evokes Hemingway's novel — *For Whom the Bell Tolls* (1940) — one of the themes of which is the idea that modern warfare annuls the value of the traditional virtues of the soldier (selflessness, courage, etc), and that only he who has the bigger guns will win, in a way that is vaguely related to Pettit's virtual selfishness about the ultimate determinative force in human action.

[225] Pettit, 1993, 2000, 2008.

[226] Pettit, 2000, p. 243.

As an illustration of such a reappraisal, Pettit takes the explanation of the persistence of the slavery system in the south of the United States. On the conventional rational choice hypothesis, the slavery system was an economic arrangement that suitably rewarded the plantation owners. Petit refers to Fogel and Engerman (1974) who argue that the purchase of a slave was generally a highly profitable investment. Slavery was not irrationally kept in existence by plantation owners who failed to perceive or were indifferent to their best economic interests. The slaveholders maintained the slavery system so long as enslaving a man was less costly than hiring him.

Evidence, however, suggests that the plantation owners did not think in economic terms. They rather stuck to what they did out of mere habit or because they conceived of those commitments, as some have suggested, in moral, quasi-religious terms. Self-regard thus had no bearing on the continuity of the slavery system. Religious thinking had. Yet self-interest may still be assigned an explanatory role in accounting for the maintenance of the slavery system. It may have had a virtual presence in the slave owner's mind. As Pettit says, "given the way it served economic self-interest, any plantation owner who began to change his behaviour would have quickly faced a serious downturn in economic fortunes and, in when confronted with that prospect, would have been inclined to return to the status quo"[227]. The slave owners would have found an economic reason to hold on to the slavery system. As a virtual reinforcement of the system, however, self-interest explains its stability[228].

3.2 The slavery system: why it could not last forever

Millar's explanation can also be similarly approached as describing a virtual scenario for the end of the slavery system. The church and the central political authorities actually played a more significant and positive role than that assigned by Millar's explanation. His account nevertheless remains of explanatory import. It shows that, even if the Church or the government had not abolished it, slavery would have inexorably ended. This is because these central authorities ordered the slave owners to act in a way that happened to suit their material interest. Had the central authorities not imposed such demands, the slave owners would not have failed to take the appropriate measures.

3.3 Residential segregation: why it will remain a fixture

Take, as a last example, the case of residential segregation. What is the process by which "the poor get separated from the rich, the less educated from the more educated, the unskilled from the skilled, the poorly dressed from the well dressed — in where they work and live and eat and play, in whom they know and whom they date and whom they go to school with"?[229] As Schelling notes, a few explanatory mechanisms immediately suggest themselves. Residential segregation can be expected

[227] Pettit, 1996, p. 70.

[228] Ultimately, the slavery system failed to maintain itself. On the rational choice approach advocated by Pettit, it means that agents had finally found no reasons, whether of the other-regarding or of the self-interested types, to continue to apply the rules of the slavery system. The institution collapsed when none of these two sorts of reasons were able to motivate the agents to continue to act in the way required to uphold the institution.

[229] Schelling, 1978, pp. 138-139.

to be the result of an "organized action"[230], for example. There are numerous historical examples of malevolent politicians who have found it in their interest to forcibly separate the inhabitants according to their racial or social identity. Strong segregationist preferences may alternatively play a crucial role in the emergence of residential segregation[231]. Because it is sustained by racial, social, or cultural prejudices, a preference for a neighbourhood mainly composed of one's fellows is certainly a *self-less* motive[232].

As plausible as these explanatory hypotheses might seem to be, Schelling proposes an altogether different possibility. His checkerboard city investigates how residential segregation could be the consequence of mild segregationist preferences among the residents. For this purpose, Schelling invites us to play a game[233] using a checkerboard, a few pennies and dimes in equal number and randomly distributed on the checkerboard. He then stipulates a few rules governing the moves of the pawns on the board: every dime wants at least half its neighbours to be dimes, and every penny wants a third of its neighbours to be pennies. They can each only move to the nearest empty square that meets their wants and they keep moving until they are satisfied. In the end of the process, the pennies and dimes form a segregated pattern.

Schelling's model describes one mechanism by which segregation can arise. It would be misleading, however, to hold that such a mechanism explains how segregation actually arises in the real world. For this purpose, economic factors and segregationist preferences are certainly more appropriate hypotheses. What Schelling's game shows however is that, even if these factors were not operative, residential segregation would still remain in place. It would be the result of mildly segregationist preferences. As long as living in a segregated neighbourhood does not leave the residents in areas that are too risky from their self-interested point of view, they will have no reason to be willing to move to more integrated environment. The checkerboard experiment shows how segregation can logically derive from an alternative set of factors than the factors that are operative in most real cases. It stresses its resilience[234].

This short overview shows that many invisible-hand explanations can be re-formulated as underlining the robustness or resilience of their *explanandum*, vindicating Pettit's theory of virtual self-interest.

---

[230] Schelling, 1978, p. 138.

[231] Cf. Emerson, Yancey, Chai, 2001.

[232] Racial, social or cultural prejudices are neither self-interested nor moral motives. They are, for this reason, like cruelty or revenge. For an analysis of these disinterested motives, see Holmes, 1990, pp. 267-286, and Elster (2009).

[233] Schelling, 1978, pp. 147-166.

[234] Pettit and Schelling are animated by the same interest in patterns that, on Schelling's words, "have the characteristic of tending to be realized in the aggregate no matter how the individuals behave who comprise the aggregates" (Schelling, 1978, p. 42). Pettit's theory of virtual interest seems to be perfectly illustrated by Schelling's checkerboard city. It exemplifies the kind of resilient pattern that the model of virtual self-interest brings to light. On close scrutiny, however, their concerns do not to strictly coincide. Schelling is interested in social patterns that frustrate agent's desires. Residential segregation *contradicts* agents' preference for mixed neighbourhoods. He describes a process at the end of which agents' preferences for an integrated environment remain unfulfilled. It is a process at the end of which the alarm bells continue to ring.
By contrast, Pettit's theory applies to patterns that equally meet a wide range of preferences, from self-less to self-interested ones. He is interested in situations where self-interested and other-regarding preferences neatly converge on the same social pattern. Unlike the residents of the checkerboard city who end up frustrated, the virtually self-interested agent ultimately finds his wishes fulfilled in all situations.

4. Possible worlds

We have so far established that explaining the resilience of a phenomenon consists in describing its pre-empted causes, that is, the factors that would determine the outcome were the actual causes not operative. Although all pre-empted causes are causes of a would-be variety, the reverse is not true: not all causes of a would-be variety are pre-empted causes. Describing how things could have happened is one thing. Another is describing how things would have happened had they not happened the way they actually happened. Although god Thor, the god of thunder is, for example, an alternative cause of segregation, it is not its pre-empted cause. A pre-empted scenario is not just an alternative scenario. It is a scenario that, on the scale of plausibility, occupies a privileged position next to the actual, true scenario. That invisible-hand explanations describe this latter kind of scenario still needs to be established.

Pettit's implicit way of dealing with this requirement can be reconstructed. It makes use of the idea of possible worlds, a notion that he has more recently invoked to elucidate the notion of resilience[235]. Pettit defines a possible world as one that displays features that do not occur in the actual world, while also sharing a few of them with the latter. A possible world resembles ours in a number of ways. Resilient phenomena can be approached as phenomena that figure in our world as well as in other possible worlds. For example, Pettit's model of virtual self-interest can be clarified as presupposing an actual world that accommodates polite, kind, other-regarding kinds of agents and possible worlds that are populated by self-interested agents. On Pettit's view, a resilient social pattern occurs in any possible worlds in which agents display behaviours that fall somewhere on a spectrum between extreme selfishness and extreme unselfishness.

The idea of possible worlds makes it possible to understand resilience in terms of degree and quality. Being shared by the actual and across many possible worlds is what characterises highly resilient phenomena. More resilient patterns occur in a greater number of possible worlds than less resilient ones. Now take two equally resilient social patterns occurring in the same number of possible worlds. As similar as they may be in regard to their *degree* of resilience, these two social patterns may nonetheless differ in regard to the *quality* of their resilience. This is because, as Pettit explains, the possible worlds in which one of them appears may connect more deeply with our interests than the possible worlds in which the other figures. One way a possible world could raise our interest — one way we may find it important to us — is the likelihood of being actualized of the possible world in which a social pattern is realized. Expanding on Pettit's view, we may here approach the various probability of realization of possible words as being linked to their resemblance to the actual world, where resemblance can be defined in terms of the overlap of the truth-values of propositions describing the circumstances relevant to the actual pattern. A pattern P' occurring in a set of possible worlds that on the whole resemble the actual world more than the set of possible worlds in which pattern P" occurs implies that P' is more qualitatively resilient than P". Note that degree and quality of resilience are two partly independent features. A certain social pattern can figure in a large number of possible worlds, and yet remain qualitatively less resilient than a scarcer seen social

[235] Pettit, 2008.

pattern. This is because the numerous worlds in which the former occur are less likely to become actual than the few worlds in which the latter figure.

These remarks about how resilience is qualitatively affected by the probability of the possible world in which a phenomenon occurs can be used to cast light on the model of virtual self-interest. Although Pettit never explicitly says it, he would certainly agree to say that self-interested deliberations and behaviours figure in possible worlds that are high in probability. As a well-known tendency that has attracted the attention of many Republicans, selfishness is actualised often enough to view it as a feature of highly probable possible worlds. If this is so, the world that invisible-hand explanations describe is not a world that is populated by strange and unlikely creatures. It is a world that is populated by credible agents.

This is all one needs to say to ensure that an invisible-hand explanation points to causes that are not just logically consistent with the actual behaviour of agents, but that are also very likely to be actualised, were the actual causes of behaviour cease to be operative. An invisible-hand explanation does not just describe *some* alternative scenario to the actual one. It describes *the* scenario that would have taken place, had things been different from what they are.

5. When the bell rings

Pettit's idea of virtual self-regard is based on three different predictions. It first predicts that agents will give a thought to their personal interest when the latter is compromised and that they need not otherwise do so. Secondly, deliberating accordingly, it is expected, will lead them to act so as to secure that interest. The theory, thirdly, anticipates *what* agents will decide to do in light of these freshly consulted self-interested considerations. It is predicted that they will reach the conclusion that they do not need to change their habitual, or culturally framed behaviour, as the latter happens to fit their self-interest very well.

Each of these predictions can however be questioned. Take, first, the idea that agents consult their self-interested preferences when their self-interest is threatened. The assumption turns out to advance two debatable empirical claims. First it confines the influence of self-interested concern to a virtual one and thus excludes the trivial possibility that agents routinely consult and compare both their selfless and their selfish preferences when they deliberate about what they should do. While rational choice theorists go wrong in seeing selfishness everywhere, Pettit goes too far in denying it any psychological reality by giving it a virtual presence only.

The first prediction secondly assumes that agents recognise when their self-interest is threatened. However cases where they fail to do so are easy to find. Also uncertain is whether they will always act in conformity with what they decided to do. Being aware of one's interest is one thing. Acting accordingly is another. Weakness of will is, among others, a well-known case where awareness does not issue in the right behaviour. Pettit will however reply that it is only on the basis of an objectivist theory of self-interest that a failure to recognise and protect it can be observed. Once someone's subjective opinion is considered to be all that is needed for self-interest to be satisfied, no failure to secure one's self-interest can be consistently stated. On this view, it never escapes agents' awareness that their interest is threatened. This is because cases where they seem to deliberate in a way that deeply compromises their self-

interest only reflect the unusual high cost at which they still are willing to secure their self-interest.

The third prediction is less easy to defend. It says that after having deliberated self-interestedly, agents are expected to remain faithful to their past choices and behaviour. Deliberating in a self-interested manner is predicted to produce no change of behaviour. As Pettit says, "self-regarding deliberation …leads most …back towards the original pattern"[236]. In the same vein, he predicts that "self-interest will kick in and stabilize the pattern"[237]. Agents will thus continue to act similarly because their current behaviour turns out to be compatible with what self-interested considerations require them to do.

The metaphor of the little posts is consistent with this prediction. The posts on each side of the straight line are designed in such a way that they *restore* the ball to its original path. Yet deliberating in light of self-interested concerns cannot be assumed to *always* have a restorative effect. There is no guarantee that taking a look at the situation from the perspective of personal advantages will lead agents to stick to their habits. There are no grounds to exclude the possibility that deliberating in self-interested terms might end up in a change of behaviour.

In fact, Pettit seems at times to recognise this possibility. He admits that sometimes "the alarms bells will ring, causing the agents to rethink and probably *reshape the project on hand*"[238]. Self-interest, he grants, may have a disruptive role in deviating people's behaviour from its usual path, giving rise to an entirely new form of behaviour. "Self interest will clock in to modify action", he says, "in cases where the actor comes to realize that the cost of failing to take interests explicitly into account has become too great"[239]. Note that, to be strictly analogous, the posts in the metaphor would have to be constructed in such a way that they would be capable of turning the ball away from, *as well as* redirecting it to, its Newtonian path. The bumpers of a pinball games i.e. the round knobs that, when hit, push the ball away in all erratic directions may here serve as an evocative metaphor.

The problem is that picturing the posts as bumpers would however impair their stabilizing effect. Pettit cannot consistently allow for the case where agents end up revising their behaviour as a result of consulting their self-interest. Allowing self-interest to have a disruptive effect on behaviour is not consistent with its recognised stabilizing power. If the model of virtual self-interest is to explain the resilience of social pattern, "reshaping the project at hand" is not a possibility that Pettit should welcome. A change of behaviour would bring an end to the social pattern that had been obtained in virtue of agents' past choices and behaviour. The model of virtual self-interest only explains the resilience of social pattern if, when they deliberate self-interestedly, agents leave the project as it is.

In sum, a reference to a virtual form of self-regard explains the resilience of certain social patterns under the following condition. The alarm bell rings, people consider whether or not what they are doing threatens their self-interest and conclude that, after all, there is no problem. They realize that the way they have been behaving does not conflict with their personal interest after all. It was a false alarm!

---

[236] Pettit, 2000, p. 243.
[237] Pettit, 2000, p. 244.
[238] Pettit, 2000, p. 240 [my emphasis].
[239] Pettit, 2000, p. 240.

Conclusion

Because Pettit is attached to the sustaining role of self-interest, he is led to disregard the various circumstances in which self-interested concerns disrupt our current behaviour. He leaves aside the case where we end up reshaping the project at hand. Such neglect comes at a price, that of restricting the scope of his theory. Pettit's theory only applies to cases where self-interested desires and self-less desires happen to sustain the very same behaviour. Only under this condition, will it explain the resilience of the social pattern obtained by virtue of this behaviour. If you let these two kinds of desire issue in two different types of behaviours, however, virtual self-interest will explain why, on the basis of some newly consulted self-interested consideration, agents would modify their current behaviour, contributing to the disappearance of a social pattern. It will explain, *pace* Pettit, the fragility of that social pattern.

IV. INVISIBLE-HAND EXPLANATIONS AS PHILOSOPHICAL EXPLANATIONS

What I find most striking about Menger's Principles of Economics is that it is not a work in empirical science at all but entirely a work of philosophy. Nancy Cartwright, 1994, p. 175.

Introduction

Carl Menger is given credit for having fully articulated the commodity theory of money[240]. The money system prevails within a group when a good is widely accepted as a medium of exchange and Menger's explanation accounts for this fact. He describes a process by which money emerges out of the barter system. When agents barter their goods, Menger says, they act in light of their immediate needs. "Each man", Menger says, "is intent to get by way of exchange just such goods as he directly needs, and to reject those of which he has no need at all, or with which he is already sufficiently provided"[241]. The barter system turns out to be an inconvenient way of acquiring goods. For an exchange to happen, it is necessary to meet someone who has what one needs and who needs what one has. This "double coincidence" as Menger calls it, is unlikely to take place as often as it ideally should[242]. As a consequence, a lot of valuable time is often required to carry out a trade and "the number of bargains actually concluded must lie within very narrow limits"[243].

Proceeding with indirect exchanges is the solution. Instead of accepting the good they presently desire to have, it is in agents' interest to provisionally accept another good with a view to exchanging it later for the good they are looking for. Cattle, skins, cubes of tea, slabs of salt, cowry shells, have, among other things, been used as such. This intermediate good must be well chosen however. It will play its function of reducing the trading costs only if it is itself a frequently traded good. In Menger's terms, the intermediate good must have a high level of "saleability"[244] (*Absatzfahigkeit*). This way it is more likely to be accepted by those who happen to own the good that one wants. Saleability is however only one condition. Others are, according to Menger, transportability, divisibility, durability. Possessing a few goods presenting these features is thus an advantage. One will not waste time looking for the person who both has what one needs and needs what one has.

Each time a good is chosen as an intermediate good, its saleability raises. But its saleability was already the reason why it was initially chosen as an intermediate good. This is therefore a self-reinforcing process. The more a given good is chosen

[240] cf. Lagerspetz, 1984, Iwai, 2001. Elements of Menger's theory can be found in the chapter VIII of the *Principles of Economics* [1871], 2007, pp. 257-261, in the *Investigations Into The Method of the Social Sciences with Special Reference to Economics* [1883], 1985, pp. 152-155, in his "On the origins of money" (1892a) and in "La monnaie mesure de valeur", an article published in French in 1892.

[241] Menger, 1892a.

[242] "Consider how seldom it is the case, that a commodity owned by somebody is of less value in use than another commodity owned by someone else! And for the latter just the opposite direction is the case. But how much more seldom does it happen that these two bodies meet!" (Menger, 1892a).

[243] Menger, 1892a.

[244] Menger defines a highly saleable good as one "possession [of which] would considerably facilitate the individual search for persons who have just the goods he needs" (Menger, 1892a).

as an intermediate good, the more its saleability increases, and the more it will henceforth be chosen as an intermediate good. Ultimately, agents accept a small number of goods (only one in some cases) as the dominant intermediate good(s)[245]. Money is born.

Menger's account of the emergence of money is controversial. Its opponents contest in particular the way it overlooks explanatory factors that allegedly play a role in the emergence of most actual money systems. Menger's explanation, it is first argued, misguidedly neglects the role of central authority. Fiat money, i.e. money created by enactments of the State, is much more widely used than commodity money, i.e. a medium of exchange that is a commercial good. Moreover, much evidence about the origin of commodity money points to state involvement. Political authorities played an important role in determining what functioned as a medium of exchange. All the evidence shows in particular that money originates in the ability of the state to impose a tax debt on its subjects[246]. It is the state that defines money as that which it accepts (e.g., a certain quantity of wheat or barley grain) in payment of taxes. *Pace* Menger, money often originated not as cost-minimizing medium of exchange, but as the unit of account in which tax liabilities were measured. Money is, in sum, a creature of the state.

A different objection to Menger's explanation is that it neglects the role of non-economic preferences. Consider, for example, the widespread use of cowry shells as a medium of exchange in many different primitive societies. If Menger's explanation were correct, the reason why they became a common medium of exchange would be that they were among the most bartered goods[247]. More plausibly, however, cowry shells were chosen as a means of exchange because they were widely praised as body ornamentations or as symbols of wealth, power, or religious beliefs. These cultural properties made cowry shells appear as a more appropriate choice than some other less meaningful but equally (or even more) saleable goods.

The way social preferences influence behaviour with respect to money is an area of great interest in economic sociology. The main idea that drives works in this field is that economic behaviour is shaped by non-economic values. Zelizer for example explores the hypotheses that "culture determines what money is, what is used as money, and how money is used"[248]. On this approach, money must not be

---

[245] Cigarettes count as money within a group if it is used as such. Analysing the conditions under which this statement is true, one can advance the following. First, the money system exists in a group *even* if the good used as money simultaneously performs another (causal) function. Cigarettes may both be used as a medium of exchange and purchased by some with a view to smoking them. In other words, for an item to fulfill the function of money, it must *not exclusively* be used as a medium of exchange. Second, the money system exists in a group even if more than one good plays the role of money. Cigarettes, shells and cattle may simultaneously be used as money within the same group without undermining the possibility of describing the group as having a system of monetized exchange. Third, the money system exists in a group even if, contrary to the majority, some of its members do not use cigarettes as a medium of exchange. To be sure, these outsiders should not outnumber those who use cigarettes as a medium of exchange. But to say that cigarettes are a *universal* medium of exchange seems to be too strong a requirement. The fact that cigarettes are *by and large* accepted as a medium of exchange suffices. This should not however allow us to claim, as Tuner, that "one can have the first instance of money, in this [Mengerian] model, with only one person intending to use the goods as money" (Turner, 1995, p. 225). Indeed it is not correct to say of a group in which only one of its members uses cowry shells as a medium of exchange that it is acquainted with the money system.

[246] Cf. Wray, 1998.

[247] Note that Menger's theory only assumes that a good is likely to be used as money when it is *believed to be* a highly traded good. It is thus not a necessary condition that the good is actually highly traded.

[248] Zelizer, 2005.

solely conceived in terms of its various economic functions, that is, in terms of its ability to serve as a medium of exchange, as a store of value and as a unit of account. It must also be approached as an efficient way of expressing one's social, moral, family or communal values. Reacting against the view that the emergence of the money system had a depersonalizing effect on human relationships, Zelizer argues that people have found creative ways not to let their life become homogenized by the rationalising effect of money. They have used various strategies in how they use money to mark the difference between, for example, friendships, kinship ties and work relationship. According to Zelizer, the fact that agents do not reason as rational calculators when they buy, sell, rent, donate their money affects the nature of money.

Menger did not directly respond to rival explanations of these kinds. He however was very familiar with the kind of conflicting methodological approaches that they exemplify[249]. Menger remained convinced of the legitimacy of his own approach and replied to his critics in the following way. Those who object to his account on the basis of its historical inaccuracy, he argues, are guilty of a confusion between various legitimate orientations in science. They more precisely fail to differentiate between, first, an "historical" orientation, i.e. one that is interested in examining the "individual" aspects of a particular money system, from a "theoretical" approach, i.e. one that will focus on the features a particular system shares with other money systems. They secondly fail to distinguish within the theoretical orientation between an "empirical" and an "exact" understanding of social phenomena[250]. Menger's account of money is an illustration of the exact orientation of science. As such, it consists in recognising the emergences of various particular money systems as exemplifications of a "strict", or unfalsifiable, law. The kind of understanding his account delivers is different from the one offered by both the historical and the empirical points of view. The reason is that it is an account that one-sidedly focuses on a few features of all particular money systems, namely those that are essential to the nature of money.

The purpose of this chapter is to critically reconstruct Menger's defence of his account of money. To bring to light the sort of explanatory power it allegedly enjoys, I will introduce the various distinctions that Menger advances in his methodological writings[251]. I first present the purpose of what he calls "theoretical economics" as opposed to that of the "historical" orientation of the discipline (section 1). A closer look at the difference between these two orientations leads to the distinction between two ways a token of a money system can present itself, namely as an instance of a "real type" or of a "strict type" (section 2.). In the next section, I examine how well economic agents lend themselves to these two modes of presentation (section 3.). I then show why Menger's explanation exemplifies the exact orientation of theoretical science, as opposed to its empirical counterpart (section 4.). In section 5., I critically examine a key element in Menger's argument, that is, the dogma of ever-constant self-interest. I review different ways of justifying

[249] Menger was the protagonist in a controversy involving Schmoller who defended an understanding of economic phenomena that recognises the non-economic values underlying economic behaviour. On this controversy, cf. Haller, 2004, Nadeau, 2003, 2005.

[250] Menger, (1996). Re-statements of Menger's argument can be found in Mäki (1990a,b, 1997), Smith (1986, 1990), Nadeau (2003, 2005), Haller (2002, 2004).

[251] The elements of this methodology are found in the *Principles of Economics* [1871](2007), in the *Investigations into the Method of the Social Sciences with Special Reference to Economics* [1883](1985), and in his *On the Origin of Money* (1892a). Menger's methodological thought is couched in non-conventional terms. Although it certainly is useful to highlight them by using more popular equivalents, as some have tried, finding these equivalents is not easy, as we shall see.

it and show that none is fully persuasive. I finally show where the value of Menger's explanation lies, in light of another Mengerian distinction between "organic" and "pragmatic" explanations. I argue in particular that, as both an organic type of explanation, Menger's account of money must be praised for its capacity to inform us about the nature of money — and a number of other institutions. It is an explanation that, in other words, offers conceptual insights in a way that is typical of philosophical explanations (section 6.)

1. The theoretical *vs.* the historical understanding of social phenomena

On Menger's view, a social structure (*Gebuilde*, formation)[252] like an institution can either be "historically" or "theoretically" understood. These "two great classes of scientific knowledge"[253] correspond to two different "points of view"[254] or to two "orientations of the striving for cognition"[255] of social phenomena. One follows either one or the other approach, depending on whether one is interested in examining the "individual" aspects of phenomena or whether, on the contrary, one aims at grasping their "general" features. The two approaches are thus not mutually exclusive and even complement each other.

Menger says that the individual aspects of a social phenomenon are known by "investigating its individual process of development"[256]. This requires that we pay attention to "the concrete relationships under which it has developed and, indeed, has become what it is, in its special quality"[257]. Menger describes the task of the historical sciences as providing an understanding of "concrete phenomena, located in space and time"[258]. Historical sciences are about the "particular", the "definite", the "individual" or, as he also says, the "specific". History deals with spatially and temporally determined social phenomena, which are "immediately experienced"[259]. Its scope, to use a philosophical term, is social phenomena taken as tokens, as non-repeatable particulars.

For example, we can investigate how the money system emerged in some particular prisoner of war camp during World War II. This investigation will have to depict the particular process by which the prisoners managed, without any means of communication and, of course, without any help from a central authority, to coordinate themselves in a way that finally led them to choose the same commodity,

[252] Whether and in what way Menger considers a social structure to be ontologically different from a natural structure are difficult questions. On the one hand, Menger's defense of the unity of science seems to support a reductivist view. Social phenomena, he argues, are sufficiently analogous to biological organisms to be insightfully described as "social organisms" (Menger, 1985). On the other hand, Menger also considers that social objects, and in particular economic objects, have a distinctive, non-reductible, mode of existence. Unlike, natural entities, he notices, they could not exist independently from the view we hold about them. To the extent that social objects (such as money) depend in crucial part on our subjective representations, a purely naturalistic description would be unable to capture their essence.

[253] Menger, 1985, p. 35.

[254] Menger, 1985, p. 35.

[255] Menger, 1985, p. 35.

[256] Menger, 1985, p. 43.

[257] Menger, 1985, p. 43.

[258] Menger, 1985, pp. 35, 37.

[259] Menger, 1985, p. 56.

namely bread, as a medium of exchange[260]. We are then dealing with what Menger calls a "definite phenomen[on]", which an historian will find relevant to relate in full details. The historical point of view enables the provision of explanations of particular facts by other particular facts. They are therefore singular explanations and singular explanations, which are not special instances of general explanations.

Any sufficiently sharp observer, however, will discover that the use of a commodity as a medium of exchange did not occur just once. Other similar special cases of commodity goods come to mind, such as the use of cowry shells by the Ojibway, the use of packs of Marlboros in Moscow in 1990, squirrel furs in Medieval Finlands, bird scalps among California tribes in XIXth Century, or the use of coconuts, ice cubes, cattle, skins, cubes of tea in others regions. From this point of view the use of bread in a certain WWII camp still presents itself as a token but as a token of a type and one's attention now switches to the latter perspective. As Menger says, "definite phenomena are repeated now with greater exactitude, now with lesser, and recur in the variation of things. We call these empirical forms types"[261]. Types are the recurrent aspects of things. It is in virtue of seeing that not only cowry shells, but also many other commodities such as bird scalps, bread and coconuts, were repeatedly purchased for the same purpose, that is, serving as media of exchanges, that money as a type can be discovered[262]. This is what distinguishes Menger's type from Weber's "ideal type", which "has the significance of a purely ideal limiting concept"[263] and hence cannot be considered as being instantiated by particular phenomena.

Once the bread is recognised as an instance of a type, one switches from an historical perspective to a theoretical one. According to Menger, the specific task of the theoretical approach is to achieve a "general"[264] understanding of the money system. Its goal is to reveal the aspects that a particular social phenomenon shares with other similar, particular social phenomena.

More than the ability to subsume a token under types is involved in a theoretical approach. The latter also requires the ability to abstract a law (or a regularity) from the particular development of the concrete social phenomena that are examined. Approaching the use of bread in a given prisoner-of-war camp *theoretically* is, as Menger would say, to recognise it as "a special case of a certain regularity (conformity to law) in the succession, or in the coexistence of phenomena… by learning to recognize in it merely the exemplification of a conformity-to-law of phenomena in general"[265]. From the theoretical perspective, the process by which bread evolved into money in Auschwitz is thus recognised as being also at work in other analogous situations.

---

[260] In *if this is a man* (1947, Chapter 8), Primo Levi describes in fine detail how Auschwitz prisoners end up using bread as a medium of exchange. I believe that such a narrative is what Menger has in mind when he speaks about an historical understanding of money.

[261] Menger, 1985, p. 36.

[262] Menger cites "the phenomena of purchase […], of supply and demand, of price, of capital, of rate of interest" as other examples of types (Menger, 1985, p. 36).

[263] Weber, 1949, p. 91, quoted by Mäki, 1997, p. 484. In view of the fact that Menger's philosophy of types predates Weber's and that the former led to one of the main achievements of micro-economics, it is somewhat surprising that the philosophy of the German sociologist has received far more attention than that of the Austrian economist and philosopher. For the beginnings of a comparison cf. von Kempski (1992).

[264] Menger, 1985, pp. 35, 35, footnotes 1, 36, 37, 39.

[265] Menger, 1985, p. 45.

On close scrutiny, more than one law or regularity can be extracted from the various particular historical instances of the emergence of commodity-money. In addition to a macroscopic law describing a regularity in the succession of two economic systems, namely the barter and the money systems, there are laws describing how agents generally behave which specify the considerations in light of which they act.

Yet it is too difficult to say which of the following two sets of apparently incompatible laws qualify as the regularities to be rightfully called for. Shall we, as Menger does, see them as exemplifying the following set of laws?

Law 1. Money stems from barter

Law 2. Agents choose highly saleable goods as media of exchange

Law 3. Agents are maximisers, that is, they choose the most efficient way of increasing the number of their possessions.

Or shall we rather treat them as special cases of this other set?

Law 1'. Money precedes barter[266].

Law 2'. Agents choose a culturally significant good as a medium of exchange.

Law 3'. Agents are satisficers. Unlike maximisers who look for the "optimal" solution, satisficers select the option that seems to address most needs [267].

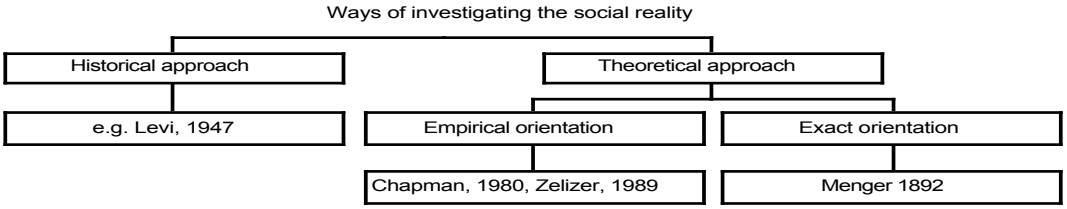
It is Menger's view that both sets of laws are circumstantially appropriate, depending on the sort of "orientation" one intends to take within the theoretical point of view[268]. Because Menger takes the "exact" orientation, his explanation of money is based on the first set of laws, which he calls "exact laws". Had he however taken the "empirical" orientation, he would have instead called for laws 1', 2' and 3'.

---

[266] It so happens that not all money systems stem from barter systems. There are indeed historically documented cases where the money system and the barter system coexist or where the former even precedes the practice of bartering goods (Cf. Chapman, 1980, p. 54).

[267] Simon, 1956.

[268] What Menger calls "the historical approach" is often confused with what he describes as "the empirical orientation of science". Mäki, 1997, Ozel, 1998, Haller, 2004, Aydinonat, 2008 conflate these two ways of investigating social reality. They should not. One way to distinguish them is to say that the first has tokens (of money) as its *explananda* and the second is about empirical types. On my reconstruction, Menger's taxonomy of explanations in social sciences distinguishes between three sorts, namely (i) the historical ones, (ii) the theoretical/ empirical ones and (iii) the theoretical/exact ones.



In light of the respective roles Menger ascribes to these two kinds of laws with regard to two ways of doing science, the next move is therefore to present this distinction in detail. As we shall see in the next three sections, exact laws differ from empirical laws in regard to: (i) the nature of their respective basic compounds, namely the "real" and the "exact" types, (ii) the way these compounds are related, namely causally and essentially, and (iii) their truth conditions, namely empirical and conceptual.

2. "Real" *vs*. "exact" types

Menger distinguishes between two ways in which the different types are discovered. To get an idea of the distinction he has in mind, consider again the use of cowry shells as money. Seeing the type in this or other similar particular phenomena is to investigate them "as these present themselves to us in their 'full empirical reality', *that is, in the totality and the whole complexity of their nature*"[269]. Seeing something in "its full empirical reality" or in "the whole complexity of its nature" is, on Menger's view, to consider both its general features and its particular features. For example, money is perceived in its full empirical reality when it is visualized as an object with a particular shape, colour and location — these are its particular and hence variable features — in addition to playing the general function of a medium of exchange. When money is examined from this perspective, its "real type", as Menger calls it, is unveiled. The real type of money therefore comprises both its universal function and the particular shapes and colours it takes on in different circumstances.

Alternatively, one can investigate the use of cowry shells as money "abstracted from [its] full empirical reality"[270]. The investigation is then focused exclusively on "a special side" of the phenomenon and leads to the discovery of its "strict type". Menger does not explicitly say what the properties of money as a strict type consist in. Yet his opinion on this matter can be partly reconstructed. These properties will be discovered, he says, when all instances of cowry-money are purified of all of their accidental properties. For example, although cowry shells are beautiful, money as a strict type does not have such property. Indeed, ugly items could just as well serve as media of exchange. It is by means of abstracting the recurrent features of all particular instances of a given social phenomenon that one obtains their strict type. Serving as a medium of exchange qualifies. It is a property shared by all particular commodities ever used as money[271].

It can however be objected that merely being a recurring aspect of the phenomenon under scrutiny is not sufficient for being the property of an exact type. Suppose, for example, that all items ever used as money happened to be beautiful. Would the strict type of money have beauty as among its properties? No, beauty would still remain an accidental property of money, one that only pertains to its "real type". Both the real and the strict types display the recurrent properties of tokens of things. Which properties pertain to strict types? The following indications can be found in Menger. First, he says that, among the recurrent properties, "the most original, the simplest and the most elementary" determine the properties of strict

[269] Menger, 1985, p. 56.
[270] Menger, 1985, p. 62.
[271] Besides being a generally accepted medium of exchange, money is commonly defined as having other functions such as a store of value, a standard of value, or a unit of account.

types. This is however rather unhelpful. Take, for example, the molecular compositions of the various things that have been used as money. They can rightly be considered as the most elementary of all their properties, yet they obviously do not pertain to the strict type of money. Take now the shapes, colours and sizes of various currencies. Aren't they the simplest of their properties? They could well be described as such and yet they do not qualify as being essential to moneyhood.

More interestingly, Menger also defines the properties of a strict type as those that express the "essence" of the things under scrutiny[272]. In the case of money, its essence is its function, viz. being a medium of exchange, the function that money performs. The properties of the strict type of money are those without which tokens of money could not perform their functions. These are the properties P_1, P_2, P_3, etc. by virtue of which something serves as a medium of exchange.

Let us consider in this respect the properties of divisibility, transportability and storability in light of this last suggestion. According to Menger, these are the properties that goods used as money often display. Are they also eligible for the properties of the strict type of money? No. Indeed, something as difficult to divide, to transport and to store as strawberries could serve the function of money, although the choice would surely not be very convenient. Being divisible, easily transportable and easily storable only make a certain good easier to use as money than another one without being truly essential to that function.

Another property that some consider as being essential to the function of money is collective acceptance[273]. Many instances of currency could not fulfil their function, it is argued, unless they were collectively accepted as such. Menger's explanation of money intends to show that this property turns out to be unessential. No one needs to agree with anyone about which of the bartered goods will serve as a medium of exchange. Each only needs to figure out which one is the most traded—that is, which has high saleability — and uses it *for himself* as a medium of exchange.

Finally Menger gives us some indication of how the distinction between strict and real types should be understood by claiming that it is not specific to the social sciences. It is also at the heart of natural sciences. According to Menger, the elements with which chemistry deals can be construed according to the same categories. For example, when it is considered in its full empirical reality, water is impure. It is composed of H_2O and of various other molecules, such as, say, HgO (mercury oxide). Water as a real type will have to take into account these other molecules. They are however of no concern to chemistry as an exact science, which exclusively deals with water as a strict type. The property of being composed of H_2O qualifies a substance as belonging to the strict type of water, unlike the property of being composed of HgO. HgO is intentionally put aside in the construction of the strict type of water because it is unessential to the nature of water. Because it is an abstraction, Menger says that pure water is "not given empirically, either *per se*, or in that ideally strict measure which the laws of chemistry presuppose"[274]. Pure water, as Menger also says, "exists only in part in our ideas"[275]. It is an abstract construction — or a partial representation — of its impure, real counterpart.

In light of these various examples and elucidations, the following can be said about the distinction between exact and real types.

---

[272] Menger, 1985, pp. 60-63.
[273] Cf. Searle (1995), Tuomela (2002).
[274] Menger, 1985, p. 85.
[275] Menger, 1985, p. 61.

First, real types should not be confused with *tokens*[276]. The particular cowry shell used as a medium of exchange, on one particular day, by a particular member of the Ojibway aboriginal people in North America is a token of money. Money as a real type is discovered by way of seeing those particular cowry shells as instances — along with cigarettes, cattle, bread, coconut, wives, cowry shells, etc. — of the sort of things that have historically fulfilled the function of money. Whereas the token exhibits only one shape, one colour and one location, money as a real type displays various kinds of shapes, colours and locations.

Second, tokens, real types and strict types are what the historical, the empirical and the exact orientations are respectively about. They also correspond to three different modes of presentation of the phenomenon that is each time under scrutiny. Both the real and exact types represent various aspects of the tokens of the same phenomenon. The real type represents these various instances more fully, because it encompasses more of the features that the tokens display[277]. The exact type is, by contrast, a much more incomplete, one-sided or partial representation of money. The distinction between real and strict types may seem to be one of degree, a real type being simply richer than its exact counterpart. However the difference is really one of essence. A strict type not only sets aside many of the properties of the phenomenon, it hold onto the most original, simple and elementary, those that constitute its essence. (Reviewing the various properties that Menger refers to in the course of his explanation, our analysis retained merely the function of serving as a medium of exchange as the only property exhibited by money as strict type. If other properties belong to money as a strict type, Menger's account does not tell us what they are).

Fourth, strict types are often interpreted as universals[278]. Because Menger believes that universals exist independently of our conceptualisations[279], he is considered as an epistemoligal realist about these. This is also what distinguishes strict types from Weber's ideal types, according to Mäki [280]. In contrast to Menger's realism, Weber's nominalism leads him to define the ideal type as "not a description of reality"[281] and as not "even a 'true' reality"[282]. And because Menger believes that universals do not exist separately from their particular exemplifications, his ontology is also classified as an immanent or Aristotelian realist one[283].

3. Economic agents

Just like money and water, an economic agent can also be approached as a token, as a real type and as a strict type. We however need to preface a discussion of Menger's view on this matter by introducing a few terms. We must first introduce the notion of economic action as one that can involve one or several economic goods.

[276] Mäki, 1997, p. 479.
[277] There are however features, such as the different smells of money tokens, that the real type of money will entirely ignore, and, hence, cannot, strictly speaking, be said to exhaustively represent tokens of money.
[278] Cf. Smith, 1990, Mäki, 1990a, b, 1997, p. 480.
[279] "The phenomena and not certain, and not their linguistic image, the concepts", he says, "are the object of theoretical research in the field of economy" Menger, 1985, p. 37. It is true that Menger also says that water as a strict type "exists only in part in our ideas" (1985, p. 61).
[280] Mäki, 1997, pp. 483-484.
[281] Weber, 1949, p. 90, quoted by Mäki, 1997, p. 483.
[282] Weber, 1949, p. 93, quoted by Mäki, 1997, p. 483.
[283] Mäki, 1990.

Examples are consuming, exchanging, selling, buying, giving, saving, accumulating, making a donation, distributing and sharing. It is Menger's view that something acquires a good-character when it possesses some features that are in the service of some end the individual has in mind[284]. An economic good is in particular something that exists given that the following are present:

1. A need on the part of some human being.

2. Properties of the object in question which render it capable of being brought into a causal connection with the satisfaction of this need.

3. Knowledge of this causal connection on the part of the person involved.

4. Command of the thing sufficient to direct it to the satisfaction of the need[285].

For example, houses satisfy my need for shelter, just as food satisfies my nutritional need and so qualify as goods. Note that Menger's definition does not admit rain as a good with regard to farmers' needs inasmuch as it does not satisfy the fourth condition, farmers being unable to make rain. Menger's definition, it is also worth stressing, equally excludes baby food from the class of goods. This is because although baby food satisfies the need of babies, the latter have no idea that the food causally satisfies their need, let alone have any command over the satisfaction of their need[286]. (Whether Menger should allow knowledge of the causal connection between the good and the fulfilment of the need, as well as command of it, not to be exclusively that of the person whose need is satisfied is an interesting question, yet one that I do not need to address for the purpose of the present argument).

Let us moreover define an economic agent simply as someone who performs an economic action. Note finally that Menger's definition of a good does not rule out the act of giving from the class of economic actions inasmuch as the recipient knows how the good he is being given satisfies his need and can control it to that effect.

Turn now to that Ojibway aboriginal person who, on one particular occasion, uses his cowry shells as a medium of exchange in the market place. He certainly is a token economic agent, since he expects his various needs to be fulfilled by means of exchanging his cowry shells for rice. He therefore also appears to whoever takes him as the object of scientific inquiry either as a strict type or as a real type. But how are these two respectively discovered?

---

[284] Cf. Zuniga, 1999.

[285] Menger, 2007, p. 52. Note that, under a liberal reading of "needs", sofas, computers and wine qualify as goods, whereas wives, employees or, for that matter, slaves do not qualify to the extent that they are not "thing" although they may often have been treated as such.

[286] One may here reply that baby food (or, for that matter, medicines delivered to unconscious patients) is not *a* good, although it certainly is good to babies. Or, one may alternatively claim that the baby cannot grasp the causal connection and has no command over it for the same reason that it is disqualified as an agent: it is not an agent because it lacks the conceptual and agentive control over events. And since it is not an agent, the only relevant agent whose needs and command come into play are the parent's. Feeding their baby is their need, the baby being fed is *their* satisfaction, met by baby food as a good, and so on. I am not sure however whether the latter suggestion accommodates the case of the unconscious patients whose care-givers may hardly been attributed a need to give him medicines.

To find out their real type, Menger argues, people will have to be considered "in their full empirical reality". It is Menger's view that, from this perspective, an economic agent exhibits all sorts of motivations which all are constitutive of his real types. He first displays economic motivations, which are defined in terms of the ability to discern and to self-interestedly pursue economic goals. "People in their economic efforts", Menger claims, "are predominantly and regularly governed by their individual interests and on the whole and regularly recognize the latter correctly"[287]. Self-interested agents act with a view to increasing the number of their possessions and choose the option that best serves this goal. But as Menger repeatedly indicates, this tendency alone does not qualify real agents "in all cases and absolutely"[288], nor "exclusively and without exception"[289]. As Menger makes clear, the real type of economic agent also exhibits various non-economic motivations:

> In the practice of economy people in fact endeavour only rarely to protect their economic interest *completely*. Many sorts of considerations, above all, indifference to economic interests of lesser signification, good will toward others, etc. cause them in their economic activity not to protect their economic interests at all in some cases, in some cases incompletely.

When he is seen though the lens of his real type, an economic agent is endowed with conflicting impulses. The influence of laws, customs, of "external compulsion" and of "public spirit" occasionally makes him act contrary to what "the desire for the most complete satisfaction of needs possible"[290] would require. And even when he pursues his material interest, "error or ignorance" often misleads him in his endeavour.

> They are, furthermore, vague and in error concerning the economic means to attain their economic goals; indeed, they are often vague and in error concerning these goals themselves. Also the economic situation, on the basis of which they develop their economic activity, is often insufficiently or incompletely known to them. Finally their economic freedom is not infrequently impaired by various kinds of relationships.[291]

At the centre of conflicting motivations, real agents end up protecting their economic interest, albeit "incompletely" as Menger specifies.

Economic agents as strict types display a much narrower range of motivations. They shall be construed as being "guided in their *economic* activity exclusively by consideration of their individual interest"[292]. Never at a loss regarding where their self-interest lies, "infallibility and omniscience"[293] are also to be added to their selected competences. The strict type of economic agent is therefore obtained by stripping its real type of all of its non-economic motivations as well as all its its misconceptions of the nature of its economic interest. The result is what Menger, echoing his methodological adversaries, ironically refers to as "the dogma of ever-constant self-interest".

[287] Menger, 1985, p. 64.
[288] Menger, 1985, p. 64.
[289] Menger, 1985, p. 64.
[290] Menger, 1985, p. 63.
[291] Menger, 1985, p. 71.
[292] Menger, 1985, p. 83.
[293] Menger, 1985, p. 84.

It follows that an economic agent who uses a good as a medium of exchange on the basis of its high saleability acts insofar as he is an instance of a strict type. His choice is exclusively guided by a thought about what it is in his material interest to do. By contrast, someone who assigns to a good its function of money because he finds it beautiful, odd, prestigious, magical or faithful to the values of the community to which she belongs, acts as an instantiation of a *real type*.

Menger concedes without hesitation that, in reality, agents rarely act according to their strict type (although they could). Most of the time, cultural and economic preferences both enter into their reasons for acting, e.g. for choosing certain goods as money. Sometimes these two kinds of preference point towards incompatible choices, but not necessarily so. Cowry shells could both be chosen because they are believed to be highly saleable and have magical power.

As Menger emphasizes, the economic agent as a strict type is an incomplete or one-sided representation of real agents. It is an agent that is deprived of many of the particular features of agents encountered in everyday life. A more complete representation of the latter would have to take into account the influence of law, custom, ignorance and error. But the influence of these non-economic considerations manifests itself only when agents are apprehended in "their full empirical reality" or as "real types".

4. The exact *vs.* the empirical orientations of science

The distinction between real and strict types grounds another important distinction within Menger's methodology. Real and strict types respectively are the elements of "empirical laws" and "exact laws". Before looking at the differences between these two, let us first focus on the broader category of laws. Menger defines a law[294] as relating different types to each other. On this view, a law consists in a statement of a connection between two types. More precisely it consists in making a connection between types qualified as either increase or decreasing. There is a law, for example, correlating "the regular drop in price of a commodity" with "the increase in supply". Another law relates "the rise in price of a commodity" to "an increase in currency". Similarly, "the lowering of the rate of interest" can be linked to a "considerable accumulation of capital"[295]. And a relationship exists between "prices" and "the increasing or decreasing of demand"[296]. Depending, however, on the sort of types that are being connected to one another, different sorts of laws will be obtained. In Menger's methodology, whereas empirical laws connect real types, exact laws deal with strict types. The reason why empirical laws are often confused with exact laws is that their linguistic formulations may very well be semantically similar. The claim that the rise in need of a certain good brings about a rise in its price can be understood either as an exact or as a strict law. However linguistically indistinguishable they may be, exact laws differ deeply from strict laws. They not only differ with respect to the sort of types that they connect, they also diverge with respect to the nature of the relation as well as their truth conditions.

[294] Or "typical relationships" as Menger also calls them.
[295] Menger, 1985, p. 36.
[296] Menger, 1985, p. 42.

Consider the law that links an increase of needs to an increase in price. Everyone knows that the reality does not always conform to it. Menger explains this discrepancy between the law and the reality it is supposed to describe in the following way[297]. Economic agents do not, in reality, behave in accordance with their strict type. In particular, distracted as they are by the pursuit of some other social or ethical goals, they do not accurately discern their economic interest or are "vague and in error"[298] about how to pursue it. They also do not enjoy economic freedom and all of these factors negatively affect the way agents protect their interest, also explaining why the price fluctuations in reality do not conform to the fluctuations stated by the exact law of supply and demand. They explain why, in Menger's words, "real prices are… more or less different from the *economic*" ones[299].

The law that predicts the use of highly saleable goods as money can be similarly regarded as subject to exceptions. Sometimes, the good that is picked up as serving the function of money does not enjoy any particularly high saleability. At other times, agents listen to their cultural preferences and choose meaningful (rather than highly saleable) goods as media of exchange.

Interpreting the law this way, that is, as being subject to exceptions, presupposes that we view the types it connects as real types. Money and economic agents are in this case conceived "in all their empirical reality". When we adopt that orientation, we do *not* isolate the "elementary" from the subsidiary (or the essential from the accidental, and the universal feature from the particular ones). We visualize money as an object with various particular shapes, colours and locations and we represent to ourselves economic agents as moved by economic *as well as* by cultural motivations.

When it is interpreted as relating real types, the law that predicts the evolution of highly saleable goods into money falls in the category of "empirical laws". Empirical laws are, according to Menger, "determined by way of observation" and "the criterion of their truth is … the empirical method"[300]. They therefore "must agree with full empirical reality from the consideration of which it was obtained"[301]. When they fail to do so, they can be regarded as false and must be revised until they match our empirical observations. Suppose that the reality would be more adequately accounted for if, in addition to highly saleable goods, culturally significant goods were recognised as the sort of goods to which economic agents assign the function of money. An empirically founded explanation of the rise of the money would then require the application of several, possibly incompatible, laws. Such an explanation will in particular have to account for cases where cultural considerations prevail over economic considerations, and *vice versa.*

Take now the second perspective and consider the types that are connected by means of a law as if they were strict ones. Many particular features of full empirical reality must now be ignored. Far from having any shapes, colours and location, bartered goods only have the property of being highly saleable goods, the consideration of which exclusively motivates economic agents as a strict type to choose them as media of exchange. Compared to this truncated reality, the law that predicts the use of highly saleable goods as money sounds true. Indeed, the side of

[297] Menger, 1985, p. 71.
[298] Menger, 1985, p. 71.
[299] Menger, 1985, p. 71 [emphasis original].
[300] Menger, 1985, p. 70.
[301] Menger, 1985, p. 70.

the reality it refers to is correctly represented by means of it. Once interpreted as being based on an exact law, Menger's explanation turns out to be a faithful depiction of the fraction of reality that it describes. Indeed, once the types to which it refers are recognised as being of the strict category, its apparent untruthfulness turns out to be a harmless one-sidedness[302].

It now should be understandable why "exact laws" are, as Menger says, "without exception"[303], or why, as he also says, "they bear within themselves the guarantee of absoluteness"[304]. Exact laws are insulated from being empirically falsified in virtue of the way they are discovered, that is, as a result of approaching social phenomena one-sidedly. Exact laws are found by *not* considering things in their full reality. They are discovered by slicing the reality into parts, by isolating the most elementary elements from the subsidiary, the essential from the contingent, the universal from the particular. "There are no strict types …when the phenomena are under consideration in the totality and the whole complexity of their nature"[305] because the totality comprises both the essential and the accidental properties. It is only once the essential properties of types are adequately hived off from their contingent properties that types perfectly repeat themselves, thus enabling the identification of exact laws.

It is Menger's view that the discovery of exact laws[306] is not specific to the social sciences. It also characterises chemistry, physics and mechanics. Consider, as Menger invites us to, the following law-like statements (or idealisations): "bodies move in vacuum, … their weights and their paths are measured exactly, … their centres of gravity are determined exactly"[307]. Reflecting on how they explain the physical reality is enlightening. Friction prevents bodies from falling according to the law of gravity. The latter appears false in regard to that noisy reality. Yet, and this is the crux of the argument, no one ever questions the law of gravity. The fact that it cannot be observed is no reason to reject it as a useful theoretical dogma[308]. The economic dogmas are, on Menger's view, not different. Thus what is perfectly acceptable in the field of physics should not be a matter of doubt in economics.

In this *reductio ab absurdo*, Menger merely assumes that idealisations in physics are acceptable without further elucidation thus begging the question about the explanatory power of idealisations in physics, chemistry and mechanics. There are various theories available but one that fits Menger's view would have to meet the two following constraints: it would first have to be compatible with scientific realism or the idea that scientists aim at a true description of reality. It would secondly have to be a unifying theory, one that applies to idealisations in explanations of the social realm as well as to explanations of the physical realm. Sorensen's view about idealisations is in this respect such a theory[309]. Yet Menger would probably not follow Sorensen when he construes idealisations as suppositions (of the sort appearing in conditional proofs) differing from assertions in virtue of various epistemological obligations (e.g. telling the truth, converging with the others) to which the supposition does not commit one. Menger is however too invested in the

[302] For a treatment of unrealistic assumptions as either harmful or harmless one-sidedness, see Mäki (2001).
[303] Menger, 1985, p. 50.
[304] Menger, 1985, p. 57.
[305] Menger, 1985, p. 57.
[306] Or of "laws of nature" as Menger alternatively calls them.
[307] Menger, 1985, p. 58.
[308] Menger, 1985, p. 58.
[309] Sorensen (2010).

truth of the dogma of ever-constant self-interest to envisage it as a premise just temporarily assumed for the sake of the argument.

A complete account of exact laws will need to specify the nature of the relation that they establish between strict types. Menger gives us some indication in this regard when he qualifies the relation holding between strict types as one of "necessity". "Strictly typical phenomena of a definite kind must always, and, indeed in consideration of our laws of thinking, simply *of necessity,* be followed by strictly typical phenomena of just as definite and different a type"[310]. No matter the amount of empirical observation, none will ever enable one to assess, let alone to disprove, the necessary relation that we mentally find between strict types. Von Mises will later say that a typical relation, that is one that's hols between two strict types, can only be discovered in an *a priori* way.

Empirical observations cannot falsify an exact law, as Menger repeatedly claims, perhaps suggesting that mistakes are not possible about them. Yet exact laws can be mistakenly enunciated, even if Menger remains awkwardly silent about the possibility. Indeed there is no reason to assume that the correct relation of necessity — that the correct typical relationship — is always asserted. Take, for example, the law-like statement according to which "economic agents use culturally significant goods as media of exchange" and suppose that its proponent meant it to be an illustration of the exact kind. The law, I here argue, is false not in virtue of some counter-examples but in virtue of its inconsistency. The law indeed contains two mutually incompatible claims. While it explicitly represents economic agents as being driven by cultural considerations, it also refers to them as instantiations of strict types and thus implicitly represents them as being exclusively driven by a self-interested goal[311]. However, self-interest and cultural considerations conflict, in the sense that cultural considerations are essentially non-self-interested. Since the best way to satisfy self-interested goal is to choose goods that are highly saleable as media of exchange, economic agents are appropriately and necessarily connected to that choice. Inconsistency is therefore one way by which an exact law can be proved false.

It should now be clear why there is no contradiction in claiming that exact laws hold without exception and yet are falsifiable. On the one hand, exact laws hold without exception because once the correct relation of necessity is discovered, it is not by pointing out alleged counter-examples that they can be proved false. Exact laws are for that matter not correctly construed as *Ceteris Paribus* laws insofar as the clause *Ceteris Paribus* is added with a view to neutralising the perturbing factors that would make the law false[312]. Although Menger's exact laws are not threatened by such interfering factors, they however remain falsifiable inasmuch as they could fail to assert the correct relation of necessity that exists between its strict types[313].

The distinction between exact laws and empirical laws delineates in turn two different orientations within the theoretical approach. Menger calls these orientations

---

310 Menger, 1985, p. 60.

311 Let us for the moment assume that an economic agent as a strict type is rightfully represented as being entirely rational and self-interested.

312 I therefore disagree with Mäki's proposal to assimilate exact laws to Armstrong's "oaken laws" that is, as laws subject to the *ceteris paribus* operator (Mäki, 1997, p. 491).

313 In order to cast light on the nature of Menger's exact law, Smith relies on the possibility, defended by Husserl and Reinach, of synthetic *a priori* truths (Smith, 1990, 1995). The truth of the proposition "an increase in needs brings about an increase in price" is similar to the truth of "red is not green", he argues, in as much as it cannot be verified by experience nor can it be assimilated to mere analytical truth. Smith's clarification is consistent with mine.

the "empirical" and the "exact" one. It should now be clear (although he himself does not make it explicit) that his explanation of the emergence of money is an application of the latter. Its merit must thus be assessed accordingly. Menger's explanation would be genuinely faulty if the laws at its core were of the empirical kind. Because however it is based on strict laws, pointing to counter-examples is thus not a way to object to it. Menger's account can*not* be criticized for neglecting factors that in reality play a role in the bringing about of the money system[314]. Yet this is precisely what his opponents mistakenly do when they charge him with ignoring the role of the state or that of non-economic preferences in the rise of money. The objection misguidedly requires of Menger's account what only an explanation based on empirical laws can offer.

As an illustration of the exact approach, Menger's explanation provides an understanding of the general features of the money system by eliciting an insight into only a "special side of human life"[315]. It assumes in particular a materially self-interested agent who is meant to one-sidedly represent real agents[316]. The latter, as Menger recognises, also have other-regarding preferences, follow customs, are motivated by public spirit, through a sense of justice or through fellow-feeling, etc. Revealing the operation of these non-economic motivations, together with the economic ones, is the task of the empirical approach and the result is a more complete representation of social reality[317]. As incomplete as the understanding of social structures of the alternative exact approach may be, it is Menger's view that it nonetheless delivers a general and "deeper" understanding. Although the exact orientation of economics deals with only "a special side of human life", it turns out to be "the most important, the economic"[318] one.

5. The dogma of ever-constant self-interest

The dogma of ever-constant self-interest plays a key role in Menger's explanation of money. It is an explanation that describes various historical cases of commodity money as resulting from agents' striving for the satisfaction of their enlightened self-interested needs. As our reconstruction of Menger's argument showed, the dogma results from approaching economic actions as strict types. It is the result of treating the perfect knowledge of one's self-interest as a recurrent and identity-constitutive property of economic actions. By contrast, public-spiritedness, altruism and love of mankind appear as accidental properties, to be thus disqualified from the range of properties that make up the strict type of economic agents.

Menger defends the possibility of treating self-interested motivations as universals of economic actions in the following way. The pursuit of self-interested

---

[314] Menger, 1985, p. 40. As Menger says, "testing the exact theory of economy by the full empirical method is simply a methodological absurdity, a failure to recognize the particular aims which the exact sciences serve." (Menger, 1985, p. 69).

[315] Menger, 1985, p. 60.

[316] Menger, 1985, p. 60.

[317] Strictly speaking, however, it would be misleading to define the empirical approach as one that describes *all* sides of the reality. Both the exact and the empirical orientations, Menger notes, imply a certain level of abstraction from full empirical reality. He says, for example, that the empirical approach to the social and political development of nations will not comprise "all sides of the life of a nation". Be that as it may, the empirical approach can be said to provide a more complete description of that reality than the one delivered by the exact approach.

[318] Menger, 1985, p. 87.

goals is, together with omniscience, the "expressions of the most original and the most general forces and impulses of human nature"[319]. They form "the most important side" of human nature. What is and what is not important may vary from one consideration to another. Importance can be, for example, a matter of having fortunate consequences. If this is so, then surely importance can be assigned to public-spiritedness, given its desirable effect in political life. Menger however offers a few reasons to believe that self-interest is more important to human nature than public-spiritedness, which are as follows.

Self-interested motivation, he says, "is by far the most common and most powerful"[320] motive. They are, as he also says, "predominantly"[321] at play. What first justifies the dogma of ever-constant self-interest is therefore its quantitative prominence. It is a more recurring feature of the full empirical reality than public-spiritedness, altruism and any of the non-economic motivations.

Results of experimental economics teach us however to be more cautious about the so-called prevalence of self-interested motives. Identifying the various and familiar situations in which the latter are thwarted in favour of various disinterested or other-regarding motives is indeed the broad purpose of that branch of economics[322]. The claim that agents are more often self-interested than other-regarding is at best only circumstantially true.

A second justification, which appears in the *Investigations*, for the abstraction involved in the ever-constant dogma of self-interest points at the desirable consequence such abstraction has on "the possibility of a rigorous economic theory"[323]. Setting aside non-economic motivations provides the welcome prospect of obtaining strict laws. "Only when we think of man as always being guided by the same motive, e.g. self-interest, in his economic actions", Menger claims, "does each action appear to be strictly determined"[324]. "This abstraction", as Menger says, "is so inevitable in determining the 'laws of phenomena' of any kind at all that the attempt to avoid it *would really nullify the possibility of determining the laws of phenomena*"[325]. Let public spirit, love of one's fellow men, custom, or some feeling for justice determine what economic actions agents perform, and no regularity will ever be found. It would, as a consequence, "collapse the basis for strict laws of economy independent of temporal and spatial conditions, and with that the basis for a science thereof"[326]. The possibility of finding out strict laws depends on the possibility of finding a typical relationship between agents' motivation and their action, and the latter requires, in turn, that disinterested motivations be set aside.

The justification, it can be replied, is circular to the extent that the activity of abstraction involved in the discovery of strict types is presented here as being governed by the purpose of providing the elements of some already stated typical relationships. Altruistic motivations are neglected, it is argued, because doing so offers the prospect of finding some regularity. We however expect altruistic motivations to be set aside because they are supposedly unessential to the performance of economic actions, independently of the law-likeness of the latter.

---

319 Menger, 1985, p. 86.
320 Menger, 1985, p. 87.
321 Menger, 1985, p. 87.
322 Cf. Güth, 1982, Fehr, 2002, Henrich *et al.* 2004.
323 Menger, 1985, p. 84.
324 Menger, 1985, p. 83.
325 Menger, 1985, p. 80.
326 Menger, 1985, p. 84.

That a certain abstraction, say of enlightened self-interest motives, would enable the discovery of a certain regularity, say that all agents are rationally self-interested, is not a reason to make such an abstraction in the first place. The justification amounts to the preposterous claim that we have the unlimited freedom to slice the reality as we like, as long as some typical relationships are subsequently obtained. The justification opens the gate for any arbitrary statements of a relationship between two variables to be absolutely true.

There are other reasons to be sceptical about the appropriateness of construing the distinction between self-interested motivation and any of the disinterested motivations listed above in terms of universal properties *vs.* particular properties.

First, a self-interested motivation would qualify as a universal feature of all economic actions if it were constitutive of all instances of motivations of economic actions. Presumably one counter-example would be disqualifying. It is easy to find instances of economic actions that are not motivated by the desire to increase possessions (assuming that economic actions are actions involving a good). Gifts, donations, sharing, throwing away and all actions by which someone dispossesses himself of at least some of his goods are obvious examples. Although it may not always be right to describe those actions as being based on the desire to decrease the number of one's possession — rage, anger or some other impulsive motives may drive someone to throw away his goods — they certainly do not seem to be performed with a view to obtaining more goods.

Some will object to these alleged counter-examples by pulling back the curtain on some hidden, self-interested goals that push us to share or give our goods. One cannot deprive oneself of a good, it is argued, unless one seeks the pleasure of giving[327]. The warm-glow effect, it can however be replied, is indeed a pleasant feeling, yet one that accrues to us only when we do not adopt it as a goal[328]. Other similar cynical ways of explaining away apparent acts of altruism as results of the desire to build an ultimately advantageous reputation have been found not credible in view of the results of numerous laboratory experiments in which the possibility of sharing a certain amount of money is given just once to an anonymous player[329].

Menger could very well reply here by excluding the acts of giving, sharing and alike from the class of economic actions. He may either reject the very possibility of those acts, or, less oddly, describe them as instances of moral actions. The move will however prove vain since it is possible to point to economic actions that do not involve any dispossession and yet are not based on enlightened self-interest. The act by which cowry shells are used as a medium of exchange is an example. The choice is a disinterested one insofar as, having no significant saleability, cowry shells are picked only due to their alleged magical powers. The various disinterested instances of economic actions given so far will cast doubt on the possibility of approaching enlightened self-interest as a universal, recurrent or essential feature of all economic actions.

It is however perfectly consistent with Menger's defence of the dogma of ever-constant self-interest to reply that these examples simply illustrate the possibility that disinterested motives are occasionally stronger than interested ones. None of these examples, Menger could reply, prove that self-interest is entirely absent from the motivational set. They are compatible with the possibility that self-interested motives

[327] Andreoni, 1989.
[328] Feinberg, 1958.
[329] Cf. Güth, 1982.

are merely less influential than disinterested ones. An action is after all not a good indicator of the ambivalence of its motives because when self-interested considerations are overridden by altruistic considerations, they give shape to an action that bears no trace of the former. We must therefore concede that enlightened self-interest lies at the core of all economic actions, sometimes in some possibly undetectable form.

There is however another reason to remain sceptical about Menger's defence of the dogma of ever-constant self-interest as a strict type of economic motivations. It concerns how the relationship between essential properties and accidental properties of the economic entities differs from that between essential and accidental properties in other sciences. The comparison reveals indeed the following two dis-analogies.

In the case of economic actions, the accidental properties, i.e. ignorance of one's self-interest and altruism, and the essential properties, i.e. omniscience and self-interest, are opposites. They are mutually exclusive with respect to a given situation but it is possible to both be self-interested with respect to some options and be other-regarding with respect with other options. Using a distinction between goals and side-constraints, Lynch and Walsh have clarified the various ways these two motives may combine, distinguishing four possibilities[330]. There is, first, the "lucrepathic profit-motive" the goal of which is to maximise, free of any moral side-constraints. There is the strong lucrephilic profit-motive the goal of which is to make a profit, albeit within some moral side-constraints. There is the weak lucrephilic profit-motive the goal of which is extra-commercial but which has profit as a side-constraint, and finally there is the lucrephobic motive the goal of which is moral virtue and which takes no profit-motive as a side-constraint. What these possibilities show is that while being compatible, self-interest and altruistic concerns are in a 'tug-of-war' type conflict. The more self-interested, the less altruistic one is (and vice versa).

Obviously, the same however cannot be said about the essential and accidental properties of water and money. Take a sample of water. It is not the case that the more molecules of H_20 are added to a sample, the less it will include of the other molecules. Consider also the essential and accidental properties of a token of money. It does not even make sense to say that the more of the former — the more it is taken as a medium of exchange — entails the less of the latter — the less colourful it will be. Oddly enough, only the essential and accidental properties of economic agents as strict types are mutually impeding.

Consider now the reason why molecules of Hg0 are set aside when scientists idealize water as H_20. Hg0 never influences the causal powers of water and so can be neglected. Similarly, the strict type of money does not include colours because any coloured item could serve as a medium of exchange. Are altruistic motivations neglected for the same reason? No, their neglect is not related to the fact that they have no influence in how economic agents behave, since Menger explicitly recognises their influence. There is therefore a disanalogy between the *rationale* governing the neglect of accidental properties from one strict type to another. In the cases of money and water, the decision to set aside HgO and the colour green, respectively, is justified on the grounds that these features are not causally relevant in explaining the core behaviour of water and money, whereas the decision to set aside altruistic motives in economic actions strictly conceived is justified on quite different grounds.

[330] Lynch and Walsh, 2004.

This is what makes it more questionable than the essential/accidental distinction in other fields.

At this point, it is still mysterious why the possibility of achieving strict laws by selecting, among the observed conflicting motivations, the exclusive influence of public spiritedness (or of love of mankind, of custom, etc.) on economic decision-making is not available. The mystery evaporates once one considers the passages in the *Investigations* where Menger fully recognises that possibility. He admits that the discovery of strict laws may not only stem from isolating the propensity to act in light of self-interest, it may also originate in isolating the influence of the ideal of justice, of altruism within the motivational set. Yet he also believes that such isolation, while perfectly legitimate, will *not* pertain to the exact orientation of economical sciences. Rather it will satisfy the purpose of the exact orientations of "social" sciences. Menger aligns the economic *vs.* social distinction with the selfish *vs.* altruistic distinction, when he says that it is a legitimate goal "to understand [the real phenomena] in an exact way… as exemplification of *social* laws, even if, as is self-evident, not as those of economy"[331]. On the one hand, he admits that the task of finding strict laws lends an equal legitimacy to the discovery of regularities within the range of customary and public-spirited actions as within the variety of materially self-interested actions. On the other hand, he also assumes without further justification that these two kinds of regularities delineate two different fields of research, namely the economical and the sociological fields. But whether the social sciences can be distinguished from the economical sciences in this manner is precisely what is contested by those who dispute his dogma of ever-constant self-interest.

Let us however admit that, much like economics, social science also has its exact orientation. However useful recognising the exact orientation of the social sciences (in contrast to the economical science) might be, it nonetheless raises another problem. It allows a token of an action to simultaneously exemplify two contradictory universal properties. Consider in this respect a given donation to the benefit of a third-world country. It is an implication of Menger's defence of the exact orientation of sciences that such action can be approached as an exemplification of two different strict types. It can first be approached as instantiating an economic action, inasmuch as it satisfies some people's need in a third-world country and will thus display only the property of self-interestedness. Alternatively, it can be viewed as instantiating a moral action, inasmuch as it is altruistically motivated and will thus only exhibit the property of altruism. It is an unfortunate implication of Menger's social ontology that it allows one and the same token to instantiate two contrary universal properties.

One solution would be to consider that when someone makes a donation, she performs two different actions, a moral one and an economic one, rather than one. Splitting the donation into two actions would solve the problem of having two universal and contrary properties instantiated by one action. But the solution is not likely to be compatible with Menger's commitment to parsimony, nor it is likely to please our intuitions that only one action has actually been performed.

To conclude, the dogma of ever-constant self-interest is justified in many ways, none of which turns out to be fully convincing. Because it is a crucial element of Menger's explanation, underlining the problems it raises casts doubt on the value of this explanation as an application of the exact orientation of science.

[331] Menger, 1985, p. 78 [Emphasis mine].

6. Organic *vs*. pragmatic explanations

Menger's explanation of the origin of money illustrates a class of explanations he calls "organic" and which he opposes to "pragmatic" ones. Although both types of explanations are explanations of the way social structures emerge, they describe a different sort of genetic process. An outcome is pragmatically explained, according to Menger, when we have "an explanation of its nature and origin from the intentions, opinions and available instrumentalities of human social unions or their rulers"[332]. Pragmatic explanations, in other words, represent their *explananda* as someone's goal. The outcome to be explained is more precisely described as the result of a common will or a legislative act. They are, in Menger's words, the "products of the agreement of members of society, or of positive legislation, results of the purposeful common activity of society thought of as a separated active subject"[333]. A social structure that is of pragmatic origin emerges either because a legislator designed it or because agents collectively agreed to act in the way that is required for that structure to arise.

An organic explanation of a social outcome is called for when a pragmatic explanation does not apply. It will be invoked when the outcome to be explained can neither be related to what agents collectively intend to produce, nor to what their ruler plan to establish[334]. Pragmatic and organic explanations perfectly maps onto the distinction previously introduced between an invisible-hand explanation and an intentional-design explanation.

Social structures that must be explained organically are more common than we may think, Menger argues. For example, "language, religion, law, even the state itself… the phenomena of markets, of competition, of money…"[335] are to be explained this way. Menger also considers these social structures as beneficial to the members of society. All the more striking is therefore the fact that no designer was involved in their emergence. The following puzzle: "*How can it be that institutions which serve the common welfare and are extremely significant for its development come into being without a common will directed toward establishing them*" is, according to Menger, "the most noteworthy problem of the social sciences"[336].

The pragmatic and the organic explanations form two competing conceptions of the origin of the money system. Menger grants the adequacy of the pragmatic explanation of fiat money. But he denies it as regards commodity money. "No evidence", he says (formulating Searle's central claim about institutional facts that they involve a function assignment), "gives credibility to the hypothesis of a designer that would have agreed to assign the function of universal medium of exchange to any of the goods that were traded"[337]. An easy way to reply to Menger is to point to the broad evidence that testifies the role of legislators in the establishment of various cases of commodity money. It is just not true that agents were never ordered to

[332] Menger, 1985, p. 145.
[333] Menger, 1985, p. 145.
[334] It will be applicable in particular when "we cannot properly speak of a purposeful activity of the community as such directed at establishing them. Nor can we speak of such activity on the part of the rulers" (Menger, 1985, p. 146).
[335] Menger, 1985, p. 146.
[336] Menger, 1985, p. 146.
[337] Menger, 1892.

assign to a specific good the function of money[338]. In fact, Menger's organicist view of institutions makes him rule out any role for the state. "Legislative compulsion", he writes "not infrequently encroaches upon [an] organic developmental process and thus accelerates or modifies the results"[339] (1985, 157). He admits that once institutions have reached a certain level of development "the purposeful encroachment of public powers on social conditions becomes more and more evident. Along with the organically created institutions there go those which are the result of purposeful social action. Institutions which came about organically find their continuation and reorganization by means of the purposeful activity of public powers applied to social aims"[340]. In sum, Menger grants the legitimacy of the pragmatist approach an explanation of the maintenance of institution is to be advanced and once their organic emergence has been properly recognized.

Yet there is a reason why Menger finds the pragmatic approach mostly unwarranted. Unlike an organic explanation, he claims, a pragmatic explanation is unable to answer the following question: "What is money?" There is something puzzling about money that only an organic explanation can elucidate. Someone unacquainted with the money system would indeed be intrigued by "the nature of those little disks or documents, which in themselves seem to serve no useful purpose, and which nevertheless, in contradiction to the rest of experience, pass from one hand to another in exchange for the most useful commodities, nay, for which every one is so eagerly bent on surrendering his wares?"[341]. Money can be mysteriously described as a good that everyone exchanges without ever making any use of. How a good can be both useless and eagerly sought after is therefore the perplexing fact, the "economic anomaly"[342], that an explanation of the money system must resolve.

A pragmatic explanation is in this respect clueless. Indeed pointing to a "legal dispensation"[343] does not cast any light on the reason why certain goods are exchanged for no apparent purpose. It is not a convincing way of solving the mystery of money because, as Menger argues, it consists in pointing to "causes lying outside the sphere of individual considerations"[344]. Why do causes of that kind explain poorly? I suppose that, according to Menger, pointing to them amounts to saying that agents counted a certain good as money because they were given the order to do so, and thus begs the question as to why the order was given in the first place. It is much more enlightening to look for causes *within* the sphere of individual considerations. It is in particular instructive to ask oneself how anyone can benefit from acquiring those apparently useless commodities. In this way, what seemed to be a useless good is now viewed as an ingenious device, one that anyone willing to avoid the cost of the barter system would adopt.

Nozick offers another explanation as to why organic explanations (or as he calls them "invisible-hand explanations") have a special explanatory import. They are "fundamental explanations", that is, "explanations of the realm in other terms"[345]. In contrast to pragmatic explanations (or as Nozick calls them, "straightforward

[338] Wray (1998) offers historical evidence of the involvement of political authority (of which the state is only one instance) in the establishment of money.
[339] Menger, 1985, p. 157.
[340] Menger 1985, pp. 157-8.
[341] Menger, 1892.
[342] Menger, 1892.
[343] Menger, 1892.
[344] Menger, 1892.
[345] Nozick, 1974, p. 19.

explanations"), organic explanations have great explanatory power because they "minimize the use of notions constituting the phenomena to be explained… they don't explain complicated patterns by including the full-blown pattern-notions as objects of people's desires or beliefs"[346]. A pragmatic explanation accounts for a pattern "in terms of the desires, wants, beliefs, and so on, of individuals, directed toward realizing the pattern"[347]. "Within such explanations", Nozick notes, "will appear descriptions of the pattern, *at least within quotation marks,* as objects of belief and desire"[348]. In more technical terms, Nozick is saying that in a pragmatic explanation part of the *explanandum* — what is explained — necessarily appears in the *explanans* — what explains, as what agents intend to bring about. The explanation therefore assumes some sort of knowledge on our part of that *explanandum*. An organic explanation makes no presupposition of that sort. It depicts agents who, while bringing about the pattern to be explained, are ignorant of what they are doing. They have not designed it, and have not even contemplated it. Anyone offering an organic explanation has therefore no other choice but to refer to the actions of agents under a description that is known by their performer and such a description will not include the pattern-notions.

Organic explanations, fundamental explanations and invisible-hand explanations, are what Williams refers to as "imaginary genealogies"[349]. Williams expands on Nozick when he finds their great explanatory power in the fact that they do not, as a constraint, represent the outcome to be explained as the content of agents' intentions:

> A story which offered a collective deliberation as the route to the outcome would presuppose what the story is supposed to explain: the people in the "earlier" situation would have already to appreciate the content of concepts such as justice and property, and their connections with reasons for actions, but it is an important aim of the story to illuminate what is involved in these things.[350]

An organic explanation operates, in sum, like a perfectly non-circular definition[351]. And what it reveals is that "money has not been generated by law"[352], that "in its origin it is a social and not a state institution"[353] and that "sanction by the

---

[346] Nozick, 1974, p. 14.
[347] Nozick, 1974, p. 14.
[348] Nozick, 1974, p. 14.
[349] Williams, 2002, pp. 31-38.
[350] Williams, 2002, p. 34.
[351] Williams goes further in the elucidation of the explanatory power of imaginary genealogies by saying that they often surprisingly represent what they explain as functional. "An imaginary genealogy", he says, "is explanatory because it represents as functional a concept, reason, motivation, or other aspect of human thought and behaviour, where that item was perhaps not previously seen as functional" (Williams, 2002, p. 34). For example, a fictitious genealogy of the state, such as the one Nozick provides, represents it as having the function to protect agents' property rights. The explanation will cast light on the institution, Williams argues, even if real states cannot be adequately accounted for in functional terms.
[352] Menger, 1985, p. 43.
[353] Menger, 1985, p. 43. Menger here classifies money as a *social* institution, as opposed to a *state* institution. Elsewhere he excludes money from the *social* realm in virtue of its relevance to the *economic* realm. The apparent inconsistency vanishes once the property of being "social" is interpreted as having two meanings. When it is opposed to a *state* institution, an institution is social to the extent that it arises *spontaneously*. When it is opposed to an *economic* institution, an institution is social in so far as it arises in virtue of cultural, moral or customary preferences.

authority of the state is a notion alien to it"[354]. An organic explanation shows, in other words, that state money is what money could be as a contingent matter, but is not essentially. Money is essentially a private institution in the following sense: first, agents can figure out by themselves, unaided in this task by some political authority, the kind of material benefit they can derive from using a commodity as a medium of exchange. Secondly, they need not be externally choreographed in order to converge on the same commodity. The high saleability of some of the bartered goods does the coordinating work.

Attempting to dispense with a designer is what Menger urges us to do when undertaking the task of explaining the origin of many social structures. Succeeding in this task, he claims, will cast light on the nature (or essence) of social entities[355]. For any given social phenomenon, it will provide an "understanding" that recognises "the reason for its existence and for its characteristic quality (the reason for its being and for its being as it is)"[356]. In other words, organic explanations have a heuristic power. Describing a social structure as if no one ever designed it reveals its true nature. Providing an explanation of how a social structure arises that successfully dispenses with a common will or a central authority will obviously show that neither a common will nor a central authority is essential to the existence of that social structure. To be sure, the explanation will look like a causal explanation, as it describes a causal process whose final stage is the money system. Yet the explanation turns out to be a philosophical explanation, one that intends to answer the following questions: What is money? What are its essential properties? Under what condition does it exist within a group? Is money real, and if so, in what sense? What function does it play? Is it a natural fact or a social fact? Is a normative fact? How do particular instances of money relate to money as a social universal? Does money exist just in virtue of people believing that it does, or is more involved? Menger's explanation, I have attempted to show, offers valuable answers to these ontological questions.

It is important to see that a common will and legislative acts may be inessential elements of the money systems, and yet play a causal role in many actual or historical money systems. Showing that the money system could very well arise without collective agreement and legal enactment does not entail that the latter have no influence in most real money systems. They may very well have such influence. But this is of no concern for the proponent of an organic explanation inasmuch as he merely intends to describe what is indispensable to the functioning of a social structure, while remaining silent about their actual functioning. In sum, organic explanation can be assimilated to a philosophical explanation whose purpose is to highlight the conditions for the existence and functioning of social phenomena.

We have so far introduced two important distinctions within Menger's methodology — the exact vs. the empirical orientation in science, on the one hand, and the organic vs. the pragmatic way of explaining institutions. A crucial and final step in the argumentation is to reveal the connection between these two distinctions.

---

[354] Menger, 1892.

[355] Mäki subscribes to a similar view when he describes invisible-hand explanations as involving a redescription of the outcome to be explained in terms of more fundamental entities, in a way that characterizes its essential nature (Mäki, 1990 (b)). My account however departs from Mäki's account in regard to the accuracy of defining invisible-hand explanations as "causal-genetic explanations"(cf. Mäki, 1990 (b), p. 329). My view is that invisible-hand explanations only have the appearance of the latter but truly are better construed as philosophical (or "essential", "conceptual", "non-causal") explanations.

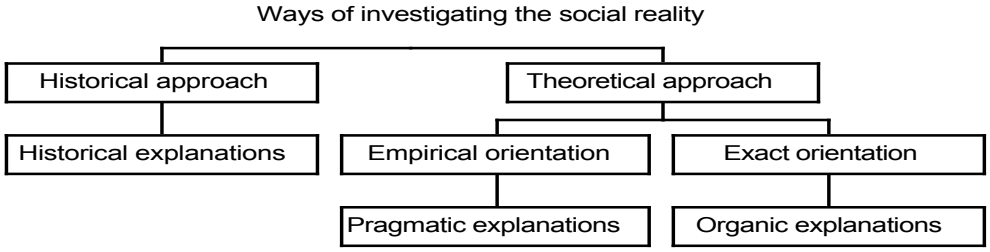
[356] Menger, 1985, p. 43.

On my view, the connection is quite straightforward: while organic explanations are relevant illustrations of the exact orientation of science, the empirical orientation is suitably exemplified by any pragmatic explanations of the institutional world. That these two distinctions map onto one another[357] is not obvious but the following considerations help to show that they do.

On the one hand, note that an organic explanation of a social structure could equally well be an illustration of the exact orientation as of the empirical orientation. There are indeed instances of money systems that arose as the result of a process involving only agents acting self-interestedly. The obligation imposed by such perspective, that is, the obligation to consider instances of money systems in their full empirical reality may be compatible with the discovery that this full empirical reality happens to only involve materially self-interested agents. In other words, while non-economic agents are often part of it, as Menger makes it clear, they need not be part of it. Sometimes real agents do act according to the dogma of ever-constant self-interest. This is why Menger's explanation of money is fundamentally ambiguous with regard to its kind. Although it is an illustration of the exact type, it may as well be an example of the empirical type. The organic *vs.* pragmatic distinction seems, in light of this consideration, not to map onto the distinction between the exact the empirical orientations.

On the other hand, it is important to note that a pragmatic explanation will never fulfil the goal of the exact orientation. To assign a causal role to legislative acts and collective agreements is to assign a role to two accidental features of the institutional realm according to Menger. It is perfectly legitimate to take these features into account, according to Menger, to the extent that money is considered in its full empirical reality. In sum, only an empirical approach to institutions could recognise their role. Or, to put it in other terms, fiat money cannot consistently be the subject matter of an explanation that is advanced from the exact perspective.

We are now in a position to offer a complete picture of Menger's taxonomy of the ways of investigating the social reality. The following tree-diagram presents the various options at hand and their relations.



[357] Menger seems to regard the two distinctions as orthogonal to one another when he allows an organic explanation to be either of the empirical-realistic type or of the exact type. "All theoretical understanding of phenomena", he says, "can be the result of a double orientation of research, the *empirical-realistic* and the exact. […] The understanding of the social phenomena which point to an "organic" origin… can also be sought in the two above orientations of research" (Menger, 1985, p. 140). I however believe that Menger considered his own organic explanation to be *necessarily* of the exact type.

Discovering that these two distinctions go hand in hand is helpful. Because we are at last able to identify the explanatory power of an explanation, such as Menger's, that is illustrative of the exact orientation of science. We can now see what it is that we know by means of an explanation that describes money, or for that matter, any institutions as the result of unintended consequences. The knowledge we end up acquiring, I have argued, is of the conceptual kind.

Conclusion

Menger's explanation of the money system is sometimes considered to be explanatorily powerless. According to its opponents, Menger's explanation unrealistically attributes too much influence to material self-interest and economic omniscience. It also allegedly overlooks the factors, i.e., the role of non-economic preference and the intervention of a central authority, which have been shown to be at play in the bringing about of most past and present money systems.

The first goal of this chapter was to reconstruct Menger's reply to this criticism. It has consisted in clarifying the goal and nature of his explanation of money in light of the various distinctions he puts forward in the *Investigations*. Here in the form of a list, these distinctions reveal the variety of ways in which the social realm can be sliced and diced:

(i) There is a distinction between two processes by which money can originate, namely, "organically" or "pragmatically".

(ii) There is a distinction between two kinds of money, namely commodity money and coined money.

(iii) There is a distinction between "economic motivations" and "non-economic motivations".

(iv) A social phenomenon can be regarded as a token, an instance of a real type, or an instance of a strict type.

(v) There is a distinction between exact laws and empirical laws.

(vi) There is a distinction between the economic realm and the social realm.

Menger's account of the money system is a one-sided explanation of the social reality and the sides of reality it specifically neglects are highlighted by these distinctions. His explanation ignores the pragmatic origin of money systems. Commodity money, rather than fiat money, is its scope. It deliberately leaves out the influence of non-economic preferences. It exclusively accounts for the elementary, original, and simple properties of money and it finally has no interest in the actual regularities in the transformation of a barter system into a money system. Once sliced in this manner, the social reality makes Menger's explanation of the rise of money true, rescuing it from the charge of inaccuracy.

I have spelled out the two sorts of confusions which Menger's critics are guilty of. Menger's *theoretical* account can first be mistaken for an *historical* one. This is what happens when, for example, it is criticized for not spelling out the specific historical

circumstances in which various particular goods have served the function of money. Even if they accept that Menger's explanation is rightly treated as a theoretical account, another confusion threatens its critics, who can fail to appreciate the *exact* orientation of this theoretical account. This is what happens when an empirical regularity is wrongly invoked as falsifying the strict laws which undergird his account.

Another goal of this chapter was to the assess Menger's argument. Its weakness, I have shown, resides in one of its constituents, namely, the dogma of ever-constant self-interest. I have reviewed several defences of this dogma and shown that none are fully convincing nor consistent with Menger's social ontology.

A final and more constructive goal of this chapter was to propose an alternative view of the value of Menger's explanation. My suggestion is to approach it as a theory whose purpose is to disclose the conditions that are essential for the existence of money (and, for that matter, for all social structures). It is a theory that intends to show in particular that perfect knowledge of one's economic interest is all that is needed for agents to act in the way that is required for the money system to emerge. Whether it is also the condition that in reality prevails over the existence of most actual money systems is another matter. If other factors turned out to be operative out there, it would merely mean that the real world is more complicated than strictly necessary.

V. SEARLE AND MENGER ON INSTITUTIONS[358]

Introduction

Two influential but conflicting approaches to social ontology are "intentionalism"[359] and the theory of spontaneous order. The two approaches advance, in particular, conceptions of institutions that seem to diverge in the following way. Intentionalists claim that institutions would not exist if we did not believe they existed. They equally stress the indispensable role of our agreement, whether in the form of a social pact or in the more casual form of a collective recognition, in the creation of institutions. While Hobbes, Locke and Kant are the oldest advocates of this conception, Searle (1995), Tuomela (2003), Gilbert (1989) and Lagerspetz (1995) are its key contemporary proponents. Advocates of the theory of spontaneous order suggest, on their part, that we treat institutions as if no designer, individual or collective, ever invented them. This alternative conception has its roots in the Scottish Enlightenment (Hume, Burke, Smith[360], Ferguson and Millar) as well as in the Austrian School of Economics (Menger, Mises, Hayek). Schelling (1978), Nozick (1974), Ullmann-Margalit (1977, 1978), Hull (1988) and Keller (1994) are some of its more recent (albeit not uncritical[361]) advocates. In a nutshell, the divergence between intentionalists and proponents of the theory of spontaneous order concerns the importance that each attributes to our consciously shared representation of institutions and of other social objects. Whereas the former claim that institutions depend essentially on our representation in order to exist, the latter pride themselves on offering a theory that advantageously dispenses with such representation.

The goal of this final chapter is to provide a clearer picture of the way these two approaches conflict. I will more specifically limit the range of the comparison by focusing on Searle and Menger's[362] respective views of institutions. There are several reasons to restrict the inquiry to these two theorists in particular. First, they offer articulated and representative versions of each of the two approaches under scrutiny. Second, Menger and Searle consider their task in similar way, that is as providing a

---

358 This chapter is partly published in the form of an article in the June 2010 issue of *Philosophy of the Social Sciences*.

359 The term is from Zaibert (2004).

360 Menger mysteriously bans Smith from the tradition of spontaneous order. According to Menger, Smith and "his followers" holds the view "that the institutions of economy are always the intended product of the common will of society as such, results of expressed agreement of members of society or of positive legislation". (Menger,1985, p. 172). "The broad realm of unintentionally created social structures", Menger therefore concluded, "remains closed to [Smith and his followers'] theoretical comprehension." (1883, p. 172). Many might however be inclined to believe that Menger's view of Smith involves a "monumental misunderstanding" of his thought, as White in his introduction to Menger (White, 1985, p. xvi) advances. Following the common view, I myself take Smith as a thinker who, *pace* Menger, did contribute to the understanding of social reality using the category of unintended consequences of individual actions.

361 Ullmann-Margalit (1978, 1994) is, for example, a supporter of the explanatory use of the notion while remaining highly critical of what she considers to be a biased "ideological use", one that she finds in Hayek's writing when he infers the goodness of an institution from its spontaneous emergence.

362 Searle, 1995, 2006, 2007, 2008, Menger, 1892, [1963/1985] 1996, [1871] 2007).

conceptual analysis of institutions. While Searle purports to reveal the "logical structure" of institutions, Menger intends to provide, as I have shown in the previous chapter, a "theoretical" understanding of institutions, one that will reveal their "nature". Moreover, we find in their writing an explanation whose initial stage is a situation in which the institution does not exist and whose final step is a situation in which it does[363]. Both Searle and Menger believe that an explanation of how institutions could emerge will cast light on its logical structure. Both also understand such explanation as different from an historically accurate account of the way institutions emerge. In their view, such explanation is different from a historically accurate account of the way institutions emerge, as neither believes in the empirical truth of the story that they tell[364]. Menger and Searle share the same goal and method, that is, of providing a conceptual analysis of institutions by means of a rational reconstruction of their establishment. Finally, the comparison between these two thinkers strongly suggests itself because both take the money system as their favourite example[365], finding in its logical structure the model of all institutions.

Searle and Menger disagree, however, about what this logical structure is. To reveal the nature and boundaries of such disagreement, I will first begin with the presentation of the three main components of Searle's theory of institutions: constitutive rules, the imposition of status functions and collective intentionality, or the fact that it is always a "we" that imposes a status function (section 1). The last element, namely collective intentionality, has no place in Menger's view of the emergence of institutions. For this reason, I will pay particular attention to an argument that Searle repeatedly offers about the essential role of collective intentionality in the creation of institutional facts (section 2). The next section shows that Menger's explanation of how money could replace the barter system represents a counter-example to Searle's defence and use of collective intentionality (section 3). I then present Menger's account of such an example and show how, on his own terms, Menger intended it to challenge the intentionalist view of institutions developed by the German historical school of which Schmoller was the leading figure (section 4). The next section deals with what could be Searle's reply to Menger, that is, the latter's inability, given his commitment to the self-interest assumption, to account for the so-called "deontic dimensions" of institutions (section 5). An exploration of a few paths towards reconciliation between the two views will finally close the comparison (section 6).

1. Searle's theory of institutional facts

In the *Construction of Social Reality*[366] Searle is concerned with the logical structure of social reality. It aims at offering an understanding of this reality that accords with the idea that we live in one world, one that is described by physicists and that is made of particles, electrons and mountains. Social reality seems not to fit

[363] Such explanation can be found in Menger (1892). Searle offers a description of the emergence of money in (1995, pp. 41-43).

[364] On the historical truth of his account, Searle writes: "I will assume this account is true, but it does not really matter much for our purposes. *I am using the account only to illustrate certain logical relations, which do not depend on its historical accuracy"* (Searle, 1995, p. 43, my emphasis).

[365] Menger offers his analysis of the money system in a paper published in 1892 entirely dedicated to the issue, as well as in chapter VIII of his *Principles of Economics* (2007).

[366] Searle, 1995, will be referred to as *CSR* hereafter.

very well in this world because it is composed of objects — money, judges, kings, marriages, etc. — that seem to be irreducible to particles and electrons. The striking difference between these two kinds of objects is how they relate to our intentional states, i.e. to our beliefs, judgments and representations. Unlike mountains and electrons, money, judges, kings and marriages would not exist if we did not represent them as existing. They are, in Searle's words, observer-dependent. The problem that Searle ultimately addresses is, therefore, how do objects that are observer-dependent fit in a world that is fundamentally composed of observer-independent objects.

Exploring this question, Searle demarcates a sub-category of social reality, namely the category of institutional facts. He considers collective intentionality to be one of the three basic "building blocks" of institutional facts[367]. The two others are the assignment of function and constitutive rules[368]. Let us present each of these components by starting with the third one.

Searle contrasts regulative rules with constitutive rules. A regulative rule merely "regulates" a behaviour that is logically independent of and prior to the rule[369]. Table manners, for examples, regulate an activity, eating, that may perfectly well be performed independently of table manners. A constitutive rule also regulates but, in addition, it creates or defines a new form of behaviour. The rules of chess create the possibility of playing chess, a possibility that did not and could not exist prior to the existence of these rules. Similarly constitutive are, by the way, the rules for promising and other speech acts as well as the rules that govern the performance of pure actions as Moya[370] defines them.

Searle further claims that constitutive rules play a central role in understanding institutional facts. They make institutional facts—such as the existence of money, of borders, of leaders, of marriages, etc.—possible. All constitutive rules, he also claims, have the following structure: "X counts as Y in C". Let us look at the variables in more detail.

The X term refers to a brute fact or object. It is a fact or an object that would be ontologically the same whether people perceived it or not. For example, cowry shells do not need to be perceived to exist. Indeed, cowry shells are things that exist independently of human intentionality. Human beings could all be eliminated without such an event having any impact on the existence of cowry shells. The Y term refers to the same thing as the X term but it refers to it under a different description, namely an institutional description. "Money," "borders," "king," "marriages," and "conferences" are examples of Y terms.

The next question is: "How can the thing that is referred as X count as the thing that is referred as a Y?" It is only if a certain function of a certain kind, namely a "status function" is assigned to the brute facts (the X) that it becomes a Y or an institutional fact. Examples of such status functions are: "To serve as a medium of exchange", "to delineate two states", or "to indicate that its bearer is a king". There seems to be a difference between status functions and the heart's function of pumping the blood. Nor is a status function seem to be of the same sort as a

[367] Searle, 1995, pp. 23-26.
[368] Searle, 1995, p. 28.
[369] Searle, 1995, pp. 27-28.
[370] See chapter I for a presentation of Moya's distinction between pure and impure actions.

screwdriver's function of loosening or tightening screws. How does Searle account for the difference between these three types of function?[371]

The distinction between agentive and non-agentive functions is what Searle uses to distinguish status functions from biological functions. Status functions are agentive, that is, they modify the range of what agents can and cannot do, unlike non-agentive functions, such as the heart's function of pumping blood, which does not modify the range of things we do. The function of pumping blood that is assigned to the heart is part of the theoretical account of the heart. By contrast, an agentive function "has to do with our immediate purposes, whether practical, gastronomic, aesthetic, educational, or whatever"[372]. We assign to screw drivers the agentive function of driving screws with the practical purpose of driving crews. We assign to wine the agentive-function of pampering our taste buds.

Money, screwdrivers and wine all serve some practical purpose. But the last two are not institutional facts. How can we account for the difference between these two and money? To do so, Searle introduces the distinction between agentive functions that are causal and agentive functions that are not causal. Status functions are non-causal agentive functions in the following sense: they are assigned to entities that are physically unrelated to the function they perform. When squirrel furs, cowry shells and cigarettes serve as media of exchange, these objects do not play this function in virtue of their intrinsic physical features. By contrast, the physical shape of screwdrivers is what enables them to perform their function. Similarly, if wine did not have a certain molecular composition, our taste buds would not be duly titillated.

The context, labelled as "C" in the formula, plays an important role in the assignment of status functions. For the same words, let us say, "Yes, I will marry you", to count as a declaration of marriage, they must be uttered in a certain context, namely one in which at least an official authority is present to register them. The same is true of the rule that ascribes to squirrel furs the function of medium of exchange which only applies in the context of Medieval Finland.

The third component of institutional reality, which will be critically discussed below, is collective intentionality. It is more precisely the collective recognition, acceptance or acknowledgement of the status-function imposed on certain entities. "We have good reasons to believe", says Searle, "that the 'counts as' locution specifies a form of collective intentionality"[373]. According to Searle, it is always a 'we' that counts an X as a Y. It is important to note that Searle does not say that individual intentionality cannot assign functions to objects[374]. An agent alone may

---

[371] One may expect Searle to say that, unlike status functions, biological functions are not observer-relative. It seems indeed that the heart has the function of pumping the blood, whether we represent this fact to ourselves this way or not. But Searle dismisses this way of distinguishing a status function from a biological function. He claims that functions of all types exist only relative to the attitudes of observers. In other words, to have a certain function is never an intrinsic feature of an object or an entity: there must be someone assigning the function of pumping blood to the heart for the latter to have the function of pumping blood. Surely, that heart pumping blood and the movement of the valves are intrinsic facts of nature. But when we assign the function of pumping blood to the heart, "we are situating these facts relative to a system of values that we hold" (Searle, 1995, p. 15). If the survival of the organism was something we valued no more than the sound the heart makes, we would assign both functions to the heart, namely that of pumping the blood and that of making noise. The fact that we deny the heart has the latter function seems to confirm Searle's idea that functions are indeed relative to some values that we hold.

[372] Searle, 1995, p. 20.

[373] Searle, 1995, p. 95.

[374] Searle, 1995, p. 39.

perfectly well assign to a log the causal agentive function of serving as a bench. Collective intentionality, he argues, is distinctively involved in the assignment of *status* functions. It implies that *an agent alone could not assign a status function to an object.* He needs to be accompanied in such assignment with the other members of his group.

When Searle claims that status functions are always assigned in virtue of collective acceptance, he excludes a mere summative account. "Status functions are not, in his view, ascribed by several "Egos." It is rather only when these "Egos" impose the function *together* that they form a genuine "we"[375]. Several agents individually assigning the same function to the same object do not act together, because to act together is not merely to pursue the same goal. Nor is it sufficient that, in addition to sharing the same goal, agents have the mutual belief that they are doing so. For example, there is a difference between two agents who are aware of the same use, that of serving as a bench, that each of them inadvertently makes of the same log, and two agents behaving similarly as part of an outdoor ballet choreography. Unlike these dancers, the first two agents merely converge on the same goal and could express what they do without reference to one another. In contrast, the dancers have a sense of doing something together[376] that can be captured by the fact that what they are individually doing, namely using a log as a bench, is derivative of what they are collectively doing, namely perform a choreography. This leads Searle to regard collective intentionality as a primitive notion, alongside individual intentionality.

Searle finds the fact that status functions are collectively — rather than individually — assigned to be one of the three crucial ingredients of social reality. Institutional facts, he claims, are those special kinds of social facts that emerge when human beings collectively award *status functions* to entities that do not have their function in virtue of their physical properties. Since having a status function defines what an institutional fact is, we can sum up Searle's view by saying that every institutional fact requires that we *collectively* count that brute fact as having the given status function.

2. Searle's argument in favour of collective intentionality

A quick reading of *CSR* and of Searle's more recent articles on the subject may convey the impression that collective intentionality is a postulated ingredient of institutional facts[377]. Searle clearly argues along this line when he says that "social facts, on my account, are *stipulatively* defined in terms of collective intentionality, and institutional facts are a special subcategory of social facts"[378]. Here collective intentionality is built into the larger category of social facts. As a sub-category of this larger category, institutional facts merely inherit this feature from their genus. Some other passages in *CSR* and elsewhere, however, support another interpretation. They

375 Cf. Searle (1990) for a presentation of this non-summative account of collective intentionality.
376 Searle, 1995, p. 24.
377 Most of his interpreters claim that Searle merely stipulates this ingredient. See, for example, Ylikoski & Mäkelä (2002, p. 469) and Hindriks (2003) who claims that "an exploration of the relation between status functions and collective acceptance reveals that Searle does not provide an argument as to why collective acceptance is required for institutional facts". *Pace* Hindriks, I will show that Searle does offer an argument in favor of the need for collective intentionality, albeit not a convincing one.
378 Searle, 1997, p. 452 [my emphasis].

support the idea that collective intentionality is the only way to explain how status functions are assigned. In these passages, Searle argues for the necessity of collective intentionality when he also points to the fact that an object does not perform its status function in virtue of its intrinsic, physical features. There, he defends the first claim by means of the second. There is an often-quoted passage in the *Construction of Social Reality* where he does just that:

> The central span on the bridge from physics to society is collective intentionality, and the decisive movement on that bridge in the creation of social reality is the collective intentional imposition of function on entities that cannot perform these functions without that imposition.[379]

Searle claims that collective intentionality is an essential feature of all institutional facts *because* of the way status functions are performed, namely *not* in virtue of the physical features of the entity that fulfils it. The argument is clearly stated in the following passage:

> I distinguish between functions performed solely in virtue of causal and other brute features of the phenomena and functions performed only by way of collective acceptance. The key element in the development of agentive functions into institutional facts comes when we collectively impose a function on a phenomenon whose physical composition is insufficient to guarantee the performance of the function, and *therefore* the function can only be performed as a matter of collective acceptance or recognition.[380]

And in the following passage:

> This step, the collective imposition of function, where the function can be performed only in virtue of collective acceptance, is a crucial element in the creation of institutional facts.[381]

And in this one:

> Since the function in question [i.e. status function] cannot be performed solely in virtue of the physical features of the X element, it requires our agreement or acceptance that it be performed.[382]

And, more recently, he claims that:

> Many functions of objects and people are performed solely in virtue of physical properties. Thus an object can perform the function of a hammer, a watch, a car, or a pen solely in virtue of its physical structure. There is, however, a fascinating class of functions where physical structure by itself is not enough, rather people have to assign a certain status to the object in question. And with that status goes a function that can only be performed in

[379] Searle, 1995, p. 41.
[380] Searle, 1995, p. 124 [my emphasis].
[381] Searle, 1995, p. 39.
[382] Searle, 1995, p. 46.

> virtue of the collective recognition and acceptance of the object or person as having that status.[383]

The most recent statement of the same argument is here:

> The nature of status functions requires that they be collectively recognized in order to do their work[384].

To paraphrase Searle, there are certain objects that do not carry out their function in virtue of their physical features, and this fact reveals the mandatory role of collective intentionality. Being physically unrelated to the function they have, the only way these objects can exercise their function is by 'we' collectively assigning them their function[385].

3. Private constitutive rules

The *collective* imposition of status functions is a requirement that seems to entail that an "I" could *not* effect the attribution of the status function that is involved each time an X is counted as a Y. But is it really so? Let us, for example, consider the possibility that John Searle assigns the full moon the function of indicating to him

[383] Searle, 2007, p. 14.

[384] Searle, 2008, p. 25.

[385] Searle's argument in favor of collective intentionality may sound mysterious in light of what he says about functions in general: that they are all observer relative. Remember that neither money nor screwdrivers, hearts, or benches execute their function solely in virtue of their intrinsic properties. An observer assigning functions need to be there, Searle claims, for any of the functions performed by these objects to exist. But if all functions are observer-relative, none can be said to carry out its function in virtue of its physical features *only*. In other words, the necessity of a collective acceptance in the special case of institutional facts cannot be explained by a feature that turns out to be true of all objects serving any function whatsoever, whether institutional or not.
As a reply, Searle could say that the apparent inconsistency results from mistakenly conflating two distinctions: (i) being observer-relative *vs.* not being observer-relative, and (ii) having the capacity to exercise a function in virtue of some of one's intrinsic features *vs.* not having the capacity to exercise a function in virtue of one's intrinsic features. These two distinctions play separate roles within Searle's argument. Being observer-relative is a feature that pertains to the *possession* of a function. By contrast, the capacity that an object has to exercise its function in virtue of its intrinsic features is a feature that Searle uses in order to assess how a given object *is able to carry out* the function that it has. That these two features are independent can be shown by the following two examples. Take a log lying in the middle of a forest and suppose that no one has ever represented that log to himself as a bench. The fact that no one has ever ascribed to it the function of serving as a bench would be a good reason, according to Searle, to say that the log lacks this function. It would not however be a reason to say that the log is unable to serve as a bench. The log is quite capable of this because it displays the required intrinsic physical shape that suffices to enable it to serve this function. Conversely, we can imagine the case of a log that has the function of serving as a bench as the result of having been assigned this function by some hikers. The log, let us further suppose, has a hidden crack so that it would fall apart if anyone tried to use it as a bench. The log is unable to serve as bench although it has been assigned this function. These two examples show that something can be able to perform a function while not having this function because no one has attributed this function to it. Conversely, something can be unable to play a certain function, although it has been ascribed that function.

that it is time he trimmed his sideburns[386]. Obviously, John Searle does not need anyone's agreement before he can assign the full moon this function as he can perfectly well be the only one on earth doing so. His personal acceptance suffices[387]. Suppose, furthermore, that not only Searle but all Californian men happen individually to give to the full moon the function of reminding each of them that they have to clip their sideburns. In this case, many people would be following the same genuine constitutive rule, but as a private rule. Would we not have obtained an institutional fact?

No, we would not. Searle could first say that many "Is" similarly counting the same X as a Y do not amount to a genuine institutional fact. The fact that all Californians follow the same rule does not amount to a genuine institutional fact. It is only when all Californians view what they are doing, when they assign the full moon the function of reminding them of the sideburn clipping task, as a "joint action," that they will be creating an institutional fact. They will have to see the designation of the function as something they *do together*, rather than as something each of them does on their own.

Responding to the full moon counter example in this way, however, amounts to granting that collective intentionality is merely a stipulated ingredient of institutional reality. It amounts to admitting that status functions could, in some cases, very well be individually assigned but that these cases must, as an *ad hoc* requirement, be excluded from the class of institutional facts.

As a counter-objection to the full moon example, the following second response seems more promising. To claim that institutional facts always involve the assignment of a status function is not to claim that every time there is a status function that is assigned, there is an institutional fact. For example, we commonly use notebooks as a reminder. To assign an object the function of reminding oneself to do something is to assign a status function. This is also a private rule since each of us adopts it regardless of what others do. But nobody would describe the practice of using notebooks as an institution. The full moon and the notebooks would be convincing counter-examples to Searle's theory if their logical structure was also found in clear examples of institutions, like money, frontiers, marriages, or elections. However, the prospect of construing the latter in terms of privately assigned status functions looks on its face quite bleak. Indeed it is hard to envisage any circumstance in which it could make sense to privately count a few bushes surrounding one's house as a border delineating the limits of one's property. Nor do we see any context in which someone may have a good reason to count some uttered words as a conference talk, regardless of what others may think about it. It seems that Searle is right to claim that status functions presuppose collective intentionality whenever an

---

[386] The example is a modified version of a counter-example that McGinn (1995) advances in his review of *The Construction of Social Reality*. In McGinn's version, "we" (rather than John Searle alone) all decide that the full moon is to have the function of indicating to John Searle that it is time he trimmed his sideburns. McGinn believes that the fact that John Searle is, in his version, the only individual who must act in conformity with the rule, rules it out as a genuine social rule. Searle could, however, reply that no matter how many agents play a role in abiding by the rule, the fact that its creation involves all of *us* makes it a genuine social fact.

[387] Maybe a less curious but similar counter-example is the case of someone who has the habit of hooking his screwdriver on his door the day before he has to visit a friend. That person is assigning his screwdriver a status function since there is nothing in the physical structure of a screwdriver that makes it a more suitable candidate than any other object to serve as a reminder. But again, and *pace* Searle, no one other than the individual who uses his screwdriver as a reminder needs to accept the screwdriver as a reminder for the latter to be able perform its status function.

institution is concerned. His only mistake, if it can be counted as such, is to neglect the cases where status functions are assigned *apart from* institutional facts.

Let us, however, consider the institution of money. As odd as it may sound, there might actually be circumstances in which it is rational to individually assign a certain brute entity, like cowry shells, the function of serving as a medium of exchange. Such circumstances have been persuasively described by Carl Menger in various places in his writings. The passages are those where he explains how money could spontaneously emerge within a group. A summarized version of such explanation is given in the introduction of the previous chapter (pp.74-76).

4. Menger's account of money

By way of this explanation, I argued in chapter 4. that Menger invites us to reflect on the minimal conditions for the money system to be created within a group. What this story shows is that there is, first of all, a point upon which both Menger and Searle agree. It is the idea that what allows a good to serve as a medium of exchange has nothing to do with its physical (intrinsic) features. As Zùniga notes, Menger's subjectivism precludes any attempt to define a social object, such as money, only in terms of its physical features:

> Menger developed a complex ontology of social objects which have a unique nature. Namely, economic objects are not merely describable by their physical properties since, for example, money is not reducible to the paper, metal, plastic, or electronic components which comprise the various kinds of money we recognize as, and indeed call, money. In fact, there is no single physical property that is common to all the members of the class of objects we call money[388].

Menger clearly shares Searle's idea that objects that are used as money do not exercise their function in virtue of their shape, weigh, colours, etc. Menger, however, would not draw the inference, as Searle does, that this reveals the mandatory role of collective acceptance. We can indeed interpret Menger's analysis as an attempt to derive the money system from the barter system without ever having to invoke any kind of collective intentionality[389]. Rather: attributing the function of being a medium of exchange to some good is something agents do as a private rule. Agents do not need to ask themselves whether others do the same. "As long as others are likely to accept my cowry shells", each agent says to himself, "it is enough for me to have cowry shells function as a medium of exchange". The mere fact that cowry shells are

[388] Zùniga, 2005, p. 137.

[389] From this perspective, Menger offers an account that is plainly illustrative of the "standard story" of money that Turner brings into play (Turner, 1995, pp. 223-29) as an alternative and more plausible account than Searle's. In line with the interpretation that I propose, Turner argues that the standard story successfully dispenses with collective intentionality, only relying on "explicit individual beliefs and intentions" (Turner, 1995, p. 225) in its description of the process resulting in the creation of the money system. But Turner and I differ on the role collective intentionality plays within Searle's theory of institutional facts. Whereas Turner interprets Searle as saying that collective intentionality is the only way to explain the normative feature of institutional facts, I read Searle as (mistakenly) considering collective intentionality to be a crucial step in the assignment of status function. As I will show in section IV, Searle does not account for the normative features of the social reality in terms of desire-independent reason for action. On my reading Searle does not call for collective intentionality in order to explain the normative dimensions of the institutional reality.

widely purchased, their high "marketability", as Menger would say, regardless of the reasons for which they have this quality, is a sufficient reason for anyone to use them as a medium of exchange. The resulting rule is both a constitutive rule and a private rule, a possibility that has escaped Searle. Collective intentionality is superfluous to the institution of money.

Interpreting Menger's account of money this way chimes with what the author intended to show even if he phrased it differently. Menger explicitly claims to have shown that money is not the product of an agreement (nor the product of a legislative act, as he also stresses). He is however aware of both the popularity and proud lineage of the opposite view:

> The idea of tracing these back to an agreement … was fairly obvious… Plato thought money was "an agreed-upon token for barter, and Aristotle said that money came about through agreement, not by nature, but by law". The jurist Paulus and with few exceptions the medieval theoreticians on coined money down to the economists of our day are of a similar opinion.[390]

In the *Origin of Money,* Menger similarly presents the view that money necessarily is the result of an agreement as the one that naturally suggests itself: "The idea which lay first to hand for an explanation of the specific function of money as a universal current medium of exchange, was to refer it to a general convention"[391]. An explanation in terms of "general convention" attributes a central role to collective acceptance in the creation of money. Menger intends to show what is wrong with this sort of explanation, namely its appeal to an agreement — to an "intentional common will"[392] — as an indispensable, preliminary step in the inception of the money system. "The task of science", he claims, "is to make us understand the institution of money by presenting the process by which, as economic culture advances, a definite item or a number of items leaves the sphere of the remaining goods and becomes money, *without express agreement of people* and without legislative acts. This addresses the question of how certain items turn into goods that are accepted by everyone in exchange for the goods offered for sale to them, even when they have no need for them[393]. The money system is instituted, on Menger's account, when most agents assign to the same good the same function, that of serving to mediate exchange. As it should now be clear, this assignment is a personal decision, one that agents take on the basis of self-interested considerations.

The individual efforts from which Menger derives the money system are the self-interested actions of agents who act purely with an eye on their material interests. As an early proponent of rational choice theory, Menger defends the idea that the best way to understand how institutions work is to model each agent in the group in which they prevail as "oriented simply to his own interest"[394].

Menger would also object to Searle's use of collective intentionality on the following ground. He would say that it leads him to overestimate the extent to which agents understand the function and the aggregated effect of their choices. Defending the view that agents accept the institutional rules just as the dancers play their part in

---

[390] Menger, 1996, p. 132.
[391] Menger, 1892a.
[392] Menger, 1996, p, 131.
[393] Menger, 1996, pp. 132-133 [my emphasis].
[394] Menger, 1996, p. 131.

the collective performance of a ballet is to misjudge the extent to which agents understand, even partially, the ins and outs of institutional reality. In other words, Searle's attachment to collective intentionality prevents him from seeing that institutions need not be (and often are not) as transparent to their participants as they are to some external, and theoretically skilled, observers. His conception would, on Menger's view, appear inconsistent with the condition of opacity that most agents find themselves in when they first create and later conform to institutional rules. The latter, Menger famously claims, should instead be described as the unintended consequences of agents' actions directed toward other ends.

Menger tells a story of the emergence of money that never appeals to any sort of collective intentionality. Surely, this story is historically implausible[395]. But it does not matter for the use we currently propose to make of it. When treated as a conceptual analysis, rather than as an explanatory model, it elucidates what sort of institutional fact money is. As the preceding section shows, its main conceptual illumination is to reveal that collective intentionality, in the form of collective acceptance, recognition, or agreement, for example, is a superfluous ingredient in this institutional fact.

To take money as a counter-example to the claim that collective intentionality is an ingredient in all institutions may appear as a provocative choice. Indeed, it challenges the collectivist position by demonstrating how on some alternative, and arguably better, account their favourite illustration of the collective basis for social institutions actually serves to subvert their case. In fact, Menger thought his analysis of money could be generalised to many other institutions:

> In the same way, it might be pointed out that other social institutions, language, law, morals, but especially numerous institutions of economy, have come into being without any express agreement, without legislative compulsion, even without any consideration of public interest, merely through the impulse of individual interests and as a result of the activation of these interests[396].

As the various examples presented in the preceding chapters testify, many have followed in his footsteps, seeking to understand other institutions as engendered by a set of actions performed by agents who are merely minding their own business. We should however be careful not to overestimate the extent to which all institutions can exhibit the logical structure he sees in the money system. There are no circumstances in which it would make sense, for instance, to privately (or unilaterally) count a range of bushes (X) as a border delineating the limits of one's property (Y). Because the bushes have the function of indicating to *others* that some territory is mine, there is no point in assigning such a function without being assured of everyone's recognition. Be that as it may, Menger's account of the money system reveals that Searle is wrong to claim that it is always a "we" that counts an X as Y in C. At least in the case of money, the aggregation of many "Is" counting the same X as a Y seems to suffice for the creation of an institution.

[395] The lack of plausibility is mainly due to its excessive reliance on rational choice assumptions, namely instrumental rationality and selfishness.
[396] Menger, 1996, p. 137.

5. The deontic dimensions of institutions

Suppose that cowry shells count as money for transactions in a group and that its members came to use cowry shells as the result of a Mengerian process. No one, in other words, ever agreed to use cowry shells as money. Nor is it the case that some influential agent stipulated that cowry shells had to be used as a means of exchange. Suppose further that someone, call him Sam, deviates from the general practice, ascribing the function of money to coconuts instead of cowry shells. Maybe Sam comes to this decision because he notices that coconuts also have a high marketability and, being temporary short of cowry shells, Sam notices that he could assign the same function to his large stock of coconuts. To be sure, Sam may well anticipate that other agents will not as easily accept his coconuts when he brings them to the market. But he may nonetheless be reasonably confident about his ability to exchange a few goods by using his coconuts as an intermediate good. After all, coconuts are frequently traded in Sam's group, so the prospect of finding a coconut consumer is bright.

It is a notable feature of this scenario that Sam cannot be rebuked for his behaviour. Nobody can oblige him not to use coconuts as an intermediate good. Under Menger's theory of money, anyone behaving like Sam cannot be censured. This is because no one seems to be under the *obligation* to use cowry shells as money in the first place.

Granted, there is a sense in which Sam *should* use cowry shells. This sense appeals to Sam's own economic interest. If Sam's goal is to increase the number of goods that he possesses, he should not deviate from what he is used to doing. Temporarily lacking cowry shells is a good reason to temporarily switch to coconuts. But Sam's long-term interest is best satisfied if he reverts to cowry shells. Yet while Sam might be accused of instrumental irrationality if he permanently uses coconuts instead, and, more precisely, of time discounting, he is, strictly speaking, under no obligation to use cowry shells, and only cowry shells, as money. The distinction is here between basic normativity, which involves just the possibility of error - and hence of being corrected, and talk of duty and obligation which involves the stronger notion of sanctions - censure, condemnation, rebukes, etc. The question is that of correctness conditions imposed by the existence of the money system—you need to use the correct goods—and that imposes by itself minimal normative constraints, beyond the self-interested instrumental sense of "should".

Searle would consider this implication of Menger's account to be at odds with any sensible understanding of money. The fact that Sam cannot be rebuked for using coconuts reveals that Menger's theory is unable to do justice to the deontic dimensions of the money system. His analysis shows how a status function can be assigned to an object, as a quasi-private rule, but it fails to show how assigning a status function creates "the deontic power (rights, duties, obligations, commitments, authorizations, requirements, permissions and privileges)" that anyone using cowry shells would not otherwise have[397]. This might be viewed as an unfortunate implication of Menger's analysis. In fact, it even puts into question whether money ever really was instituted through the process he describes. Any theoretical account of money should indeed accommodate the fact that anyone living in a group where cowry shells are assigned the function of money *has an obligation* to use them as

[397] Searle, 1995, pp. 121-127.

money under all circumstances[398]. To put it in Searle's terms, anyone must assign to Xs — *and to Xs only* — the function that is associated with the Y term.

The deontic dimensions of constitutive rules are deeply implicated in the reasons for which agents follow them. Sam and his peers should know that he is under the obligation to count cowry shells as money, *whether he wants to or not*. The obligation Sam is under is unaffected by what he may or may not desire to do. Searle claims that when agents recognise that a certain object has a status function, it gives them a "desire-independent reason" for acting in a certain way[399]. It is characteristic of institutional forms of powers, rights, obligations, duties that they create reasons for action that "are independent of what you or I anyone else is otherwise inclined to do"[400].

In Menger's view, however, motivations cannot be desire-independent in this way. There is no institution, or no enduring one, whose rules do not fundamentally suit the agent's desires, her self-interested inclination to act in accordance with them. This is why the only normativity that Menger allows is a normativity that pertains to our instrumental rationality. The only motivation that is compulsory is the motivation to act in accordance with what one's interest tells one to do, and self-interest, as should now be clear, prescribes the use of *any* object as money, as long as the object is a highly bartered good.

Does this mean that Menger would deny the "rights, duties, obligations, commitments, authorizations, requirements, permissions and privileges" that Searle recognises in all institutional facts? No, it does not. But he would regard this deontic apparatus as essentially related to the intervention of a legislative authority, one that, as a matter of general policy, may have prescribed the exclusive use of a certain commodity as money[401]. If you erase such legislative authority and its codification, then Menger would probably claim that you will have removed the basis for the deontic system that Searle finds essential to all institutional facts.

More crucially, Menger believed he had shown the futility of the intervention of a central authority in the creation of institutions. At least his explanation of the emergence of the money system conspicuously dispenses with it. But if so, then the various deontic dimensions that Searle identifies appear as even more useless. Why do agents need to be externally obligated to observe the rule of an institution which, being only attentive to their material self-interest, they will end up following in any case? An official enactment is unnecessary, in Menger's view, only adding stability to an institution that already works by itself. Nor do agents need to have their choices made to converge on the same good by an external coordinator. The latter seems useless in light of the self-reinforcing nature of the process that leads from the barter system to the money system, leading Menger to conclude that "money is not an invention of the state. It is not the product of a legislative act"[402]. This is not, as it now should be clear, an empirical claim but a conceptual point.

---

398 Even Lewis (whose theory of convention has been criticised for not being able to account for the normative dimension of institutions) defines money as a medium of exchange that agents must accept "without question" (Lewis, 1969, p. 24).

399 The notion of a "desire-independent reason" to follow the constitutive rules of institutional reality is explicated at length in *Rationality in Action* (2003) as well as in many recent articles.

400 Searle, 1995, p. 70.

401 Menger [1871], 2007, p. 262.

402 Menger [1871] 2007, p. 261.

The state is, however, not entirely deprived of a role in the full establishment of the money system. Its import is, as we have seen, that of making "legally binding"[403] the use of the unique good that it has authorized, converting it into a "universal substitute in exchange". For this reason, the state deserves credit. For although "it is not responsible for the existence of the money-character of the good, it is responsible for a significant improvement of its money-character"[404].

While agreeing on the mandatory effect of codification, Menger does not share Searle's idea about the way political authorities obligate us. He explicitly defines the obligation to use the official currency as an obligation that is based on a cost-benefit calculation. More precisely, the obligation to use the official currency is there when the risk of being sanctioned outweighs the benefits of using an alternative currency. Comparing the ruler to the "victor" and the agents being ruled to the "vanquished," he believes that fear plays a prominent role in our obedience to the institution of laws:

> The man in power or intellectually superior can set certain limits to the discretion of the weak men subject to him or of those mentally inferior. The victor can set certain limits for the vanquished. He can impose on them certain rules for their action to which they have to submit, without considering their free conviction: from fear.[405]

Searle flatly rejects this view, claiming that the fears of possible sanctions cannot be the reason why we have the obligation to follow a constitutive rule. The deontic phenomena, he claims, "are not reducible to something more primitive and simple. We cannot analyze or eliminate them in favour of dispositions to behave or fears of negative consequences of not doing something. Famously, Hume and many other have tried to make such eliminations, but without success"[406].

Searle insists in various places on the inadequacy of the view that recognises utility maximization as the only sort of motivations behind the acknowledgement of institutional facts:

> It is tempting to some to think that there must be some rational basis for such acknowledgment, that the participants derive some game theoretical advantage or get on a higher indifference curve, or some such, but the remarkable feature of institutional structures is that people continue to acknowledge and cooperate in many of them even when it is by no means obviously to their advantage to do so.[407]

Searle even argues (or merely stipulates) that, were agents motivated to follow a constitutive rule as a matter of inclination, no institutional fact would be created as a result[408]. To be a component of institutional reality, the constitutive rule must be followed by agents who, unlike animals, do not consult their "inclination" as to whether they should follow it or not. This is why ants are not able to impose status functions, although they are admittedly able to "mark their territory by means of

[403] Menger [1871] 2007, p. 262.
[404] Menger [1871] 2007, p. 262.
[405] Menger, 1996, p. 217.
[406] Searle, 1995, p. 70.
[407] Searle, 1995, p. 92.
[408] Searle, 1995, p. 71.

chemical signals that do not block others by sheer physical insurmountability," as McGinn recalls[409] with the intention of casting doubt on the sophisticated intentionality that Searle requires from agents involved in the assignment of status functions. But ants remain unable to create in full an institutional fact such as a border. This is because, whereas ants have an *inclination* not to cross the chemical line, agents have a *desire-independent reason* not to cross a national border without any license.

The creation of desire-independent reason stems, according to Searle, from an initial commitment or, as he more recently says, as "something like a promise"[410]. Searle explicitly recognises the kinship between his view and social contract theory[411]. The acceptance of a status function that is involved in the creation and maintenance of institutional facts is an act of commitment in the following sense. When one recognises that X counts as Y, one creates a desire-independent reason to act in a certain way, regardless of one's desire to act in this way or not. This commitment to act in a certain way irrespective of one's desire to act this way clearly has no natural place in Menger's analysis of institutions. Unlike Searle, Menger does not see the point of untying the reason to follow the institutional rules from one's personal interest to do so.

6. A path of reconciliation

So far we have shown that Menger and Searle defend opposing accounts of institution. We have in particular highlighted the divergent role they attribute to collective intentionality and to desire-independent reasons for action in the creation and maintenance of institutions. Menger's account, we have also stressed, is consequently unable to account for the (non-instrumental) normative dimensions of institutions. We have related this shortcoming to the significance he sees in matching the *rationale* of institutional rules with agents' individual interests.

Having shed some light on the points of contention between Menger and Searle regarding the logical structure of institutions, I wish now to indicate how their two conflicting approaches could, to some extent, be reconciled.

The elements of this reconciliation can be found in a distinction that Searle makes between formal and informal institutions[412]. Informal institutions are, according to Searle, institutions that are not codified into explicit laws, though they

---

[409] McGinn 1995, p. 39.

[410] Searle, 2008, p. 29. In (2003), Searle expands on the idea in the following way. I have to now, say, pay my beer, he claims, in virtue of some *prior* and *free* action through which I created for myself a desire-independent reason to P. Searle thus assigns a crucial role to a certain class of actions, performed in the past, and that have a binding effect on me now. Here are a number of questions that will be asked regarding these actions. What sort of actions can create desire-independent reason to act in a certain way? What makes these actions binding? Are they binding in virtue of the fact that they are based on a moral principle? Are there any circumstances in which I can perform these actions without binding myself? Why is it important for Searle to ensure that these actions are freely performed? How free am I really to act in a way that will have binding effect on how I will have to act later on? Can I spend my life avoiding performance of any of these actions?

[411] "[Social contract theorists] thought that there is no way we could have a system of political obligations, and indeed, no way we could have a political society, without something like a promise, an original promise, that would create the deontic system necessary to maintain political reality" (Searle, 2008, p. 29).

[412] Searle, 1995, p. 53.

could be. Examples are cocktail parties, friendships, and dates. Property, marriage, and money are examples of formal institutions[413]. Formal institutions are not logically different from informal institutions in the sense that both involve assigning a status function to an entity. However, only in the case of a formal institution is this assignment "a matter of policy"[414].

Why do we codify? When do we need to attribute to the continuing practice of assigning a status function the status of an explicit law? Searle answers the question in the following way. We would codify cocktail parties, he says, if "it mattered tremendously whether or not something was really a cocktail party or only a tea party"[415]. Codification comes into play, in other words, when it really matters whether a given particular X must be counted as a Y rather than as a Y'[416]. Take the example of money. The reason why we codify money is, following Searle, because it matters whether a given piece of paper with a portrait of Giaccometti (X) is a 100.- Swiss Francs (Y) or a forgery (Y'). We codify the money system, in brief, to prevent counterfeit money. Preventing counterfeit cocktail parties surely is, by contrast, not as important. Nor should anybody be reprimanded for occasionally confusing a cocktail party with a tea party. In fact, there is a significant difference between the two confusions. Whereas a token of counterfeit money remains a token of counterfeit money even if everyone using it believes it is real money, a party whose organisers initially intended it to be a cocktail party but which everyone mistakes for a tea party thereby becomes a tea party.

Now Menger's invisible-hand explanation of money can be approached in light of the distinction between formal and informal institutions. What his explanation reveals is that, although it often has a formal existence, the money system could very well emerge in a group in an informal or un-codified form. It shows that, after all, money is not different from cocktail parties. In the case of commodity money, something only needs to be widely viewed as money to be money and the possibility of counterfeit money is just not there. It is only when a legislator makes it a matter of policy that all Xs must count as Y that it matters tremendously whether a particular X is a Y or a Y'.

In fact, Searle accepts this view. He grants that the current codified status of the money system in most countries is a contingent state of affairs and that the money system could alternatively take the form of an informal institution. In a passage (that could have been written by Menger himself), Searle imagines

> a system of exchange where objects are held for the purposes of barter, even though the people holding those objects may have no interest in them or use for them, as such. A similar situation existed, by the way, in the former Soviet Union at the time of its collapse. In Moscow, in 1990 and 1991, packs of Marlboro cigarettes had attained the status of a kind of currency. People would accept payment in Marlboros, even though they did not themselves smoke. The combination of paper and tobacco already had an agentive function, named by the word "cigarettes", and on top of that function was imposed the agentive function named by "medium of exchange".[417]

[413] Searle, 1995, p. 88.
[414] Searle, 2008, p. 25.
[415] Searle, 1995, p. 88.
[416] It is left unspecified what considerations in particular it is appropriate to weigh when assessing whether it tremendously matters or not if an X is a Y or a Y'.
[417] Searle, 1995, pp. 42-43.

Using Marlboro packs as a medium of exchange never was a matter of general policy in Moscow. Still, the cigarettes had "attained the status of a kind of currency", as Searle acknowledges, giving credence to the possibility that money could have an informal existence. In a manner reminiscent of Menger's view, Searle is also ready to admit of the deontic dimensions only in the case of formal institutions leaving informal ones in the hands of the subjective and externally uncontrolled representations of the agents. He says that it is only "where the imposition of status function according to the formula [x counts as y in c] becomes *a matter of general policy*, [that] the formula acquires a normative status"[418]. Agents have the right and obligation to count a cowry shell as money only when a central authority makes it a matter of obligatory rule. Searle correlatively restricts the very possibility of constitutive rules to the cases where the assignment of function is "a matter of policy":

> When the practice of imposing a status function becomes regularized and established, then it becomes a constitutive rule. If the tribe makes it a *matter of policy* that [someone] is the leader because he has such and such features and that any successor as leader must have these features, then they have established a *constitutive rule* of leadership[419].

On such view, constitutive rules are, strictly speaking, codified, i.e. officially authorized, rules only. If this is so, Searle might even be willing to claim that it is only when it is a matter of policy that someone is the leader that agents have a desire-independent reason to obey him. As long as the practice of counting someone as the chief is not centrally established, agents will only listen to what their desires dictate them to do when evaluating whether to follow her directives. As long as the practice of counting someone as the chief remains un-codified, such directives turns out to be mere baseless imperative utterances[420].

It turns out that both Searle and Menger believe that the imposition of function does not entail any of the deontic powers listed above. When a rule remains un-codified, no obligation, rights, and duties of any sort are involved. Again, it does not mean that it is impossible to be mistaken, to take a cocktail party for a tea party or to take coconuts, rather than cowry shells, as the widely accepted medium of exchange. But nobody will be appropriately blamed for doing so. What is more, if the mistake becomes widespread enough, it just ceases to be a mistake altogether. The special role that Searle attributes to official codification in the creation of constitutive rules allows his theory to accommodate Menger's insight.

Conclusion

"Status function", Searle repeatedly claims, "is the glue that holds human societies together"[421]. This is because status functions allegedly are imposed on entities in such a way as to make collective intentionality compulsory. Assigning a status function is like dancing the tango: it is the sort of activity that you cannot do

[418] Searle, 1995, p. 48, my emphasis.
[419] Searle, 2008, p. 25 [my emphasis].
[420] Searle is unclear on this particular issue, so it remains a speculative interpretation of that part of his theory.
[421] See, for example, Searle, 2005, p. 9.

alone. In fact, it is not even enough to be two agents having the same mutual belief about what the other is doing. For, crucially, you must have a sense of doing something together.

This chapter started by arguing against this view. It brought into play a few cases —the full moon, the notebooks, and Menger's story about the emergence of the money system — in which status functions are a matter of individual assignment. In these examples, agents impose a function on an object whose physical features are not sufficient to guarantee the performance of the function. But agents impose this status function as part of a private rule, whether for the sake of their own comfort, or to be good-looking (e.g., being shaved), or for the sake of their own material conditions (e.g., adding to one's possessions). In view of these counter-examples, it seems fair to say that status functions are not the compelling bonding device that Searle believes they are and that what constitutes the "cement of society" remains an open question.

Menger's conception of institutions persuasively reveals that a constitutive rule could very well be a private rule. It is however a conception that is not immune to criticism. I have shown that it is, in particular, unable to account for the (non instrumental) normative dimensions of institutions and I have related this shortcoming to the unfortunate significance he sees in grounding the *rationale* of institutional rules in the self-interest of agents. The idea of desire-independent reasons that Searle uses to elucidate such normative dimensions seems, in this respect, a promising explanatory tool.

Having shown why they are conflicting views, I have finally indicated how they can also be partly reconciled. I have suggested in particular that the scope of Searle's theory starts where the scope of Menger's theory ends. Whereas the latter is interested in the money system while it still is in its informal, officially un-codified form, Searle, on his part, focuses on the money system as it matures and takes on a codified form becoming, as he says, a matter of general policy[422].

[422] I here assume that an institution exists first informally and that it subsequently evolves into a formal system. But the fate of institutions may vary, returning to a lowly informal state after having been crowned with the laurels of formal codification.

CONCLUSION

The invisible hand is a two-dimensional theory. It is, on the one hand, well known as a normative theory, one that politicians, economists and philosophers invoke to defend the privatisation of public goods, free entrepreneurship, and restricting state intervention in the economy. It is also, on the other hand, an explanatory theory, one that conveys a particular understanding of social outcomes, independently of the virtues of those outcomes. The present account explores what I believe to be the lesser known of the two faces of the invisible hand, that is, its explanatory value. Whether institutions ought to be left to the guidance of the invisible hand or not is a question I do not address.

1. An anti-We-ness theory

Invisible-hand explanations stand in stark contrast to intentional-design explanations. The latter account for the social realm as the planned effect of some individual, or collective, intelligent design. The former is commonly defined as an explanation that consists in re-describing social outcomes as unintended consequences. In chapter I, I explored various ways of interpreting this claim. My goal was to spell out the various puzzles one needs to solve in order to determine whether a given consequence is unintended or not. Cumulatively, these puzzles contribute to weakening the standard account. It is not by means of the distinctions between unintended and intended consequences, or so I concluded, that invisible-hand explanations can consensually be differentiated from intentional-design explanations.

Defining the scope of the invisible hand in terms of unintended consequences is not only an ineffective classificatory criterion. It is also, I next argued, unnecessarily restrictive. Agents need not lack the individual intention to bring about a certain social outcome in order to qualify as agents who are led by the invisible hand. It suffices that they not be acting jointly towards that end. In other words, the invisible hand can accommodate individual intentionality, whilst ruling out only collective intentionality as being superfluous to explaining the social realm.

On the proposed account, the invisible hand is an anti-we-ness theory and seeing it this way is illuminating in many respects. First, it makes it a notion that is easier to recognise and to apply. The reason is that on this view we no longer have to resolve the various puzzles encountered when dealing with the question whether a given social outcome is or is not intended. Seeing the invisible hand as an anti-we-ness theory also makes it more attuned to the current debates in social ontology. It replaces a quite outdated definition, handed down from the 18th century, which presents the invisible hand as a theory challenging social pact theories of social reality. As it turns out, social pact theories have now lost much of their appeal. Few social scientists, if any, endorse them nowadays as a compelling approach in the field of social ontology. Construing the theory as an anti-we-ness theory, by contrast, sets it in opposition to a widespread contemporary social theory. There have been in the past two decades a growing interest in collective intentionality, a notion which has been proved very helpful to the clarification of joint actions, of shared cognitions (e.g. beliefs, opinions, judgements) and collective affective states. The invisible hand

will, I hope, appear as a much more intriguing theory once it is defined as disputing many central applications of this more current and broadly shared view.

The present dissertation also aims to throw new light on the invisible hand by construing it as a type of philosophical inquiry. Let us summarise the main insights gleaned in this regard.

2. The explanatory power of the invisible hand

A widespread conception of the invisible hand I have critically looked at in chapter III sees its merit in its ability to explain, not how social patterns emerge, but why and to what extent they are resilient. Pettit in particular has defended this position by offering an unconventional conception of a key component of invisible-hand explanations, namely rational choice agency. In response to those who find this theory of agency empirically inaccurate, and, therefore of little explanatory value, he proposes to conceive it as representing, not the actual, but only the virtual cause of behaviour. Pettit happily concedes that the emergence and persistence of a social outcome is not obtained in virtue of our often dormant rational choice agency. What is explained is instead its resilience, or why this social outcome is likely to remain in place even if and when circumstances turn agents' motivational make-up toward more self-interested concerns. I then pointed out the difficulties this way of conceiving the explanatory power of invisible-hand raises. I have shown in particular that it hastily assumes that agents will, even through the prism of purely self-interested considerations, continue to act just as they had been acting from the perspective of other-regarding thoughts. It ignores the equally possible alternative scenario that they will rather reconsider the wisdom of their past action and adopt a new course of action, consequently giving rise to entirely new social patterns.

A still further way of looking at the explanatory significance of the invisible hand is to take it as exemplifying what Menger calls "the exact orientation of science". In chapter IV, I lay out this hypothesis, taking as an example Menger's explanation of money. I show that, just like any other invisible-hand explanation, it attributes no role to any political authority or any common will, and therefore flies in the face of historical truth. In light of Menger's methodological writings, I argue that its fictitious perspective is not to be seen as an some sort of falsehoods, distortions, lies, or game of make-believe but is better construed as mere incompleteness. Menger provides an incomplete understanding of the money system to the extent that it sheds light on its essential aspects only, filtering out the accidental elements.

Even though it seems informative primarily about the origin of money, it actually explicates what money is. It casts light on the features that are constitutive of the money system by showing how it emerges within a group whose members have no idea what money might be and yet end up being the proud owners and authors of a monetary system. Menger tells the story of how, given a few assumptions, the self-interested behaviour of individual trader-barterers will, with no explicit coordination, inexorably converge towards using the same commodity as a medium of exchange, thereby producing an implicit and rudimentary monetary system. In this way he shows how money is ultimately nothing more than a system of private rules which individual agents find it in their interest to obey and by which a good comes to serve as a medium of exchange. Thus the institution comes into existence before agents come to recognise it for what it is, and its mutually beneficial nature.

Two important observations need to be made here. First, the project of describing an outcome as an unintended consequence, a constraint that previously appeared unwieldy (chapter I) and unnecessarily restrictive (chapter II), is given a positive appraisal in this chapter IV. Whoever endorses it as a restriction on how explanations need to be shaped will end up proffering accounts endowed with deep explanatory import. This is because what will be presented is no less than an explanation that does not presuppose what it is supposed to explain. If you attempt to explain the emergence of money within a group without assuming that its members have any idea of what money is, you will present all the conditions under which money exists. You will end up listing all and only the various ingredients — the physical and the mental features — failing which money could not exist.

The explanation can be regarded as a philosophical one because it uses a priori conjecture rather than empirical observation and testing. The explanation would rightly be regarded as a philosophical one if it involved the forming of concepts (rather than the forming of institutions). Following Williams[423], I have assumed that invisible-hand explanations do account for the emergence of social and institutional concepts. In other words, the explanation has philosophical import to the extent that it can be regarded as one that accounts for how we acquire the concept of money, of states, etc. Yet it is contentious claim. Whether agents at the end of an invisible-hand process do master, for example, the concept of money or the concept of a minimal state can be disputed. The sceptics will argue that an invisible-hand explanation merely elucidates how, say, money comes into existence, and once it has been noticed we can afterward form a concept for it. Yet the difference between using something as money and noticing that money exist within one's group is not easy to determine. I will come back to this question latter.

The invisible-hand explanations certainly have the confusing appearance of causal explanations. Do they not describe a process made of several sequential stages causally following one another, and the closing phase of which is the occurrence of a social outcome — an institution, a social pattern, a state of affairs? Invisible-hand explanations have the appearance of genealogies (of "genetic explanations" as Hempel would call them) by means of which the lineage of states, of money, of residential segregation, of slavery are uncovered. It is therefore easy to misconstrue them, i.e., to take them to be causal explanations, and to miss their explanatory value. The main criticism they have repeatedly drawn, i.e. the fact that they are of negligible explanatory value, I argued, stems from this misconception. Not getting the facts right is an unforgivable trait in causal explanations. Conjectural accounts that need not be so much factual as fact-adjacent, are however legitimate tools for philosophers. They enable them to sort out the essential from the accidental amongst the various features that compose reality and thus to discover the fundamental conditions for the existence of things.

3. A possible misunderstanding

I would like to dispel a possible misunderstanding of the central claim of the dissertation[424]. I may appear to be defending the view that invisible-hand explanations are valuable insofar as they rule out the entities — rulers, collective

[423] See footnote 339.

[424] Markus Haller has drawn my attention to it.

agreements, moral norms, collective intentionality — that cannot figure in an explanation that conforms to methodological individualism (hereafter MI). This is not the case. MI, let us recall, denies any collectivist entities, such as the State, the Army, the Bourgeoisie, the Clergy, etc causal power, unless these terms are only a short-hand way of referring to rulers, soldiers, bourgeois and priests, etc. Because, on this view, invisible-hand explanations are exemplifications of methodological individualism, their value is regarded as derived from the soundness of the methodological principles that they engagingly serve to illustrate[425]. Invisible-hand explanations are, in other words, illustrative defences of MI.

Against this interpretation, I shall point out that, however valuable MI is, invisible-hand accounts do not in fact well illustrate MI, let alone defend it. Two reasons in particular lead me to this view.

First, the entities that invisible-hand accounts dispense with—legislators, collective agreements and moral norms—are not the sort of collective entities which MI is forbidden from appealing to. They can appear within an MI explanation without violating its principles. Rulers, for example, can be invoked as agents who make decrees and the latter are not only individual actions, they also shapes agents' behaviour by affecting their reasons for actions, doing so in a way that does not commit us to assuming a mysteriously unmediated influence of decrees on the behaviour of individuals[426]. Collective agreements can similarly figure in an explanation that conforms to MI, insofar as they are defined as agreement between individuals who mutually promise to comply with some agreed-upon rules of action[427]. Finally, there is nothing collectivist in the claim that agents act with an eye on moral concerns. When they reject the needs of legislators, of collective agreements, the friends of the invisible hand are not picking a fight with methodological collectivists. They are taking up arms against those — such as Plato, Hobbes, Carlyle, Rousseau, among others — who believe that legislators and collective agreements are essential building blocks of social reality and the representatives of this illustrious school of thought can very well endorse methodological individualism.

Secondly, methodological individualism is a theory that imposes constraints on causal explanations. It is a theory that recommends the formulation of agent-based mechanisms, that is, mechanisms that show how macro-scale states are obtained in virtue of the mental states and actions of individuals. On my account, invisible-hand explanations are not illustrations of methodological individualism, since, for this to be the case, they would have to be causal explanations. But they are not.

4. Further inquiries

[425] In fact, I argue the opposite in chapter II, where I claim that it is only when tied to a non-reductive account of collective intentionality that invisible-hand explanations can be distinguished from intentional-design accounts.

[426] Of course, a legislator often does not act alone when he invents and enacts rules. But we do not need to assume anything that is not reducible to individual mental processes in order to explain how he and his associates work together.

[427] I have however claimed in chapter II that invisible-hand accounts can be distinguished from intentional-design accounts only if we-ness is not conceived in a reductive fashion.

In the last part of the dissertation, I took a critical perspective on the invisible hand. Once identified as a form of philosophical inquiry, showing what is valuable and what is misguided with it was my next goal. Its strength, I have shown, is to ingeniously challenge Searle's view of collective intentionality as a building block of social reality. By considering various examples, I indeed showed that status functions can be *individually* assigned thus ruling out the need for a special kind of collective intentionality, such as the one Searle puts forward, namely collective acceptance. The argument I put forward, I should however emphasise, does not show that collective intentionality is a superfluous ingredient of social reality. It only defeats a particular argument, namely Searle's, in its favour.

Still, as I argued, the invisible hand way of accounting for the normative dimensions of social reality is much less convincing. The theory of the invisible hand seeks to relate our obedience to social norms to some beneficial consequences. It regards the social norms as what enables the satisfaction of some pre-existing desires that we all have for, say, security, reputation, or the possession of material goods. Social norms are regularities of behaviour we ought to conform to given the self-interested desires that we have. The norms that fail to meet these selfish needs, it is sometimes added, are ultimately eliminated from the social realm.

As some have already argued[428], however, the nature of social obligation is not adequately understood in this manner. The apparent problem with this view is that it has the following unacceptable implication: if you lack the desires that social norms will enable you to satisfy, if following the norm is not personally advantageous to you, you will thereby have no obligation to conform to them. And yet, far from always finding it in our interest to follow social norms, we quite often obey them despite powerful, contrary motivations. Many people send Christmas cards in December only reluctantly, secretly wishing nobody did[429]. Moreover, we do not relieve people of their obligation to relinquish a bus or train seat to the elderly just because they happen to have no interest in their reputation or in the pleasures of gratitude. The extent to which you value your future reputation or the extent to which you are afraid of sanctions are just not relevant considerations. You must follow the social norms regardless.

Even more broadly conceived, as including non-selfish preferences among the desires we attempt to satisfy[430], the invisible-hand view of social norms will always link the reason for which we follow them, or should follow them, to some desire of the agent. It is this assumption that Searle objects to by means of the notion of desire-independent reasons. Because desire satisfaction does not play such a prominent role in norm-regulated actions, dispensing with desires altogether, as Searle suggests, seems to be a felicitous way of resolving the issue. I suggested how promising this approach is. It is however at the current stage in need of further development and this line of inquiry would therefore constitute a fruitful extension of the project of this dissertation[431].

---

[428] Cf. Elster, 1985, 1988, 1989, 1993, Anderson, 2000.

[429] Schelling offers it as an example of an undesirable equilibrium (Schelling, 1978, p. 31).

[430] The problem with the self-interest theory is *not* merely that it has yet to identify the sort of desire(s) that social norms will properly satisfy. The problem is that it regards the normativity of social norms, or the fact that we ought to follow them, as being contingent upon their capacity to satisfy some individual, self-related desires.

[431] There are other attempts to approach social norms in a way that dispenses with self-centred pro-attitudes. The Kantian idea of "commitment", the possibility of acting on the basis of others' desires, as proposed by Sen (Sen, 2007) and Kaufmann's idea of shared desire that no one individually holds

Also crucial to the difference between Searle and Menger is the place each of their theories provides for the knowledge of concepts. Both Menger and Searle subscribe to the notion that we cannot have money (or marriage, a State, banks, universities, etc.) without the related concepts. But it is worth noting that Searle, unlike Menger, offers an account of the creation of social reality that presupposes that agents already have the corresponding concepts. By contrast, it is possible to read Menger's account as one that explains how a situation comes about which allows us to form the concept of money. That crucial point has only been alluded to in passing so far and is in need of further investigation. There are many ways of understanding this difference between Searle and Menger. An easy but superficial way to formulate it is to say that whereas Searle is interested in the creation of *institutional facts*, Menger explores the creation of *institutions*. Thus, institutional facts presuppose the existence of institutions — the fact that coconuts count as money presupposes the existence of the money system. The creation of the fact that coconuts count as money can only take place within a society whose members know what money is. Searle does not account for the formation of the various concepts upon which institutional facts depend in the first place. By contrast, how concepts emerge is precisely what an invisible-hand explanation directly answers as it outlines a process whose initial state is one in which agents are totally unfamiliar with the concepts of money, of minimal states, and of marriage. To be sure, there are institutions (e.g. banks) that depend on previous institutions (e.g. money), but the dependency stops at some point. It stops when the very first institution upon which the others depend is created just as the result of a decision to assign a status function on an object whose physical features play no role in the function it plays.

A more interesting explanation of the difference between Menger and Searle will recall the place that speech acts, or a special kind of speech act, namely declarations (enactments), play within Searle's social ontology[432]. The reason why Searle assumes that agents already have the concept of money when they impose on cigarettes the function of a medium of exchange is that he believes that this imposition needs to take the form of a declaration. Status functions, Searle argues, are created by declarations, that is, by means of linguistic facts that belong to the class of speech acts. It is manifestly true in the case of marriage — you do not get married unless someone verbally pronounces you married — but it is also true of all institutional facts such as the fact that cowry shells are counted as money, or buildings counted as prisons, universities or police stations. Each time, a declaration of the form: "We (or I) make it the case by Declaration that the Y status function exists"[433] is involved (a declaration pronounced by someone to whom the ability to make such declaration is bequeathed in the same linguistically articulated fashion, i.e. speech acts, by individuals on whom the ability to give to someone the ability to make such a statement is conferred, and so on[434]). Searle is concerned with giving an account of the genesis of the fully-fledged institutions, that is institutions with true normative 'bite', and for institutions to attain that level of maturity, Searle's idea of explicit formulation of the assigning of status functions came into play. i.e. fully fledged normatively robust institutions require us to have the relevant institutional concepts.

---

in particular (Kaufmann 2005, see also Schmid, 2005). It would be interesting to explore how these various positions differ from one another.

[432] Searle makes this claim very central to his view in Searle (2010). Cf. also Smith, 1983 and the appendix, p. 127.

[433] Searle, 2011, p. 93.

[434] Whether this potential regress is problematic I do not know.

It is here that Searle and the friends of the invisible hand come to clash. We have, on the one hand, the latter who strongly recommend anyone advancing an explanation of an institutional fact not to refer to it in the course of the explanation with the corresponding concept, on pain of circularity. On the other hand, we have Searle who believes it to be just impossible to come up with an explanation that avoids using those concepts. The fact that both Searle and the friends of the invisible hand take their explanation to be obviously illuminating of institutional reality makes their disagreement all the more worthy of additional inquiry.

Finally, there is an interesting question that I do not address in the dissertation and it is how narrative accounts, such as invisible-hand explanations, constitute philosophical explanations. The question is more precisely whether the diachronic aspect of the analysis adds anything other than entertainment value to the conceptual elucidation that is being offered. After all, most philosophical analyses that are proposed of States, money and other institutions do not take the storytelling outlook of invisible-hand explanations. As compared to these more conventional conceptual analyses, does the invisible hand narration offer equivalent, inferior or even superior results? Dealing with these questions is another avenue along which the project of the dissertation could be further explored.

APPENDIX I[435]

Already in the *Construction of Social Reality*[436], Searle advanced a claim he himself described as "radical". The claim is that institutional facts are language-dependent facts. In his recent *Making of the Social World*[437], Searle makes that "very strong theoretical claim"[438] even more central to his social ontology. Such insistence deserves scrutiny.

Money, kings, universities and frontiers would not exist, Searle first convincingly argues, if we did not believe that they existed. They are what he calls "observer-dependent entities", drawing our attention to the crucial difference between brute facts such as the fact that Mount Blanc has snow on its summit, on the one hand, and those facts that are dependent on human agreements, such as cocktail parties, football games and marriages, on the other hand. It is also Searle's view that a social fact requires much more than our thoughts and representations: the thoughts, he quite surprisingly adds, must be expressed in language[439].

The claim is rather strange. It just does not seem to be the case that we take pains to utter our intimate thoughts each time we represent to ourselves a piece of paper as a one dollar bill, each time we think of a building as a university or each time we see in the grouping of a few human beings an auction, a marriage or a cocktail party. For that reason, needing to verbally articulate in each case the "X as Y" kind of representations that are involved is unsurprisingly a contested claim of Searle's social ontology (cf. Moural, 2008, McGinn 2011, Little, 2011, Hindriks, *forthcoming*). In his review of the *MSW,* McGinn expresses such a widespread skepticism in the following way: "if we all regard certain things as money and use them that way", he reasonably asks, "isn't that enough to make those things money, without our having to say it out loud (or by sign language or some such)? To be sure, we can't have marriage without the concept of marriage; but once we have the concept and collectively ascribe it to pairs of people, don't we have all we need for the institution of marriage to exist—what need is there for uttering the word?"[440].

It is far from obvious why the thoughts that are partly constitutive of institutional facts need to be linguistically articulated. However counter intuitive the claim about the necessity of language may be, it is not a bold assertion. Searle does offer various arguments in favor of that claim—three on my count—which can be found in the *CSR* as well as in the recent *MSW*. While the first three sections critically review these arguments (sections 1-3), the last two sections expound on the third and, as I intend to show, most persuasive of these arguments (sections 4 & 5), with the help of a few additional distinctions that I borrow from Reinach (1913) and Moya (1990). Hopefully, Searle's contentious claim will on the proposed fleshing out sound more acceptable.

---

435 The appendix is a shortened version of Tieffenbach (2011).
436 Searle, 1995, hereafter referred to as CSR.
437 Searle, 2010, hereafter referred to as MSW.
438 MSW, p. 11.
439 McGinn (2011).
440 McGinn, 2011.

1. An *eo ipso* linguistic move

As it has previously been noticed (cf. chapter V, pp. 103-106, for a presentation of Searle's social ontology), Searle makes status functions the ultimate foundation of all other building blocks of institutional reality. It is from the way status functions are assigned that many other essential components — such as collective intentionality and deontic powers[441] — are derived[442]. Unsurprisingly, it is also on the same foundation that Searle bases his case for the necessity of language. All institutional facts involves a move, i.e. the move from the X to Y in the formula X counts as Y in C, and that move cannot take place without language.

Searle offers various arguments for the need for language within the performance of status functions — various ways of explaining how language is necessary to the move from the brute facts to institutional facts. The following three explanations can in particular be extracted from his writings:

1. The move, he first argues, "is *eo ipso* a linguistic move, even in cases that apparently have nothing to do with language" (CSR, p. 63). On such an *eo ipso* view[443], the thoughts which partly constitute institutional facts are not the sort of thoughts that one could have independently of language.

2. The move, he alternatively argues, involves language inasmuch as a linguistic representation is the only way to give visibility to the move from X to Y — the visibility that is needed for the performance of status functions.

3. The move, Searle finally argues, exists in virtue of a speech act inasmuch as that is the only way by which the changes in reality that are involved in the creation of institutional facts can be accounted for.

These three explanations will be spelled out in the three next sections. The last two sections propose a few refinements on the third one.

1. An eo ipso move

Searle gives two conditions for a fact to be language dependent (CSR, p. 63):

(i) The move from X to Y is constituted by thought.

(ii) The thought is language dependent.

[441] On a less conventional reading of Searle (1995) which I propose in section 6 of chapter V, deontic powers arise as the result of codification, rather than as the result of status functions assignment.

[442] I will presently say nothing about these two components as I believe that they are not germane to the subject at hand.

[443] Cf. Moural (2008) for a discussion of that particular argument.

We cannot form the thought without any language in which the thought is expressed or described. As Searle says, I need to have "some words or word-like elements to think the thoughts" (CSR, p. 63). Words are thus needed in order to be able to have the thoughts that are involved in the act of counting Barack Obama as the president of the United States. Just as the words "king", "money" and "university" are required in order to have the thoughts that something is the king, is money, is a university. The thinkable, as Searle also says, cannot in this case be "detachable from the speakable or writable expression" (CSR, p. 68). Searle metaphorically refers to words as the "vehicle" of the thought (CSR, p. 73), as "something to think with" which "we have to have" (CSR, p. 73).

In this first argument, words are needed in so far as they are "linguistic symbols" (rather than, says, communicative devices). So the symbolizing power of words is the feature in virtue of which words are essential to the existence of kings, banks, and money. Searle recognizes in all linguistic symbols three essential features. They first have to "symbolize something beyond themselves" (CSR, p. 66). Defining linguistic symbols this way is hardly helpful since it is redundant but the idea is roughly the following. A linguistic symbol means, refers to, represents, expresses, or is about something else[444]. Searle adds as a second condition that a symbol symbolizes "by convention" (CSR, p. 67). Everybody's agreement, or at least everybody's happy or reluctant recognition, of the symbolizing power of a representative device is required for the latter to count as a linguistic symbol. And thirdly, linguistic symbols are public symbols, so that a road sign that is invisible to many is not a linguistic symbol.

Note that the definition is at that point incomplete. Take, for example, the fox that symbolizes cunning or the scythe that symbolizes death. On Searle's definition, these symbols meet the three criteria and yet shouldn't they be considered as *pictorial* symbols? It seems indeed that we should restrict the class of linguistic symbols to include those that are verbally articulated only. As we shall see later it is by conflating these two categories of symbols that Searle unpersuasively proves the necessity of language.

The crucial question is: Why is the thought involved the move from X to Y language dependent? Searle provides several different answers to this question. He first argues that the thought is too complex to be held without words and gives the following thought as an example: "Her mortgage is largely paid off, but the recent decline in interest rates may make it desirable for her to refinance to lower her payments and to take out cash." (Searle, 2011). No doubt that this particular thought is impossible to have in a pre-linguistic form. But not all thoughts expressing the imposition of status function are of that complexity. As McGinn recalls, biologists have shown that ants are able to "mark their territory by means of chemical signals that do not block others by sheer physical insurmountability"[445]. If ants are able to create an institutional fact such as a frontier, the representation involved in the creation of such fact certainly does not require language.

Searle points to a second reason why a thought is language dependent. The dependency may derive from the fact that the thought itself refers to a language. A

[444] The only and remarkable exception to that feature seemed to be the self-referential word "WORD". However, it is however only when the word "WORD" refers to itself that it is such an exception. But are there any circumstances, beside maybe in contemporary art, where "WORD" has such self-referential meaning?

[445] McGinn 1995, p. 39.

case in point is the following thought: "Mt Everest has snow and ice at the summit, is a sentence of English" (CSR, p. 60). The idea is compelling enough but does not cast light on institutional facts whose creation involves thoughts that typically do not refer to any language. Take, for example, the thought: "That yellow line is a frontier". It could be held by someone who does not also think that "that yellow line is a frontier" is an English sentence" and only the latter could not be held without words. Now it is true that the thought "that yellow line is a frontier" is an English sentence" is involved in the move from an X, i.e. a brute sequence of sounds, to a Y, i.e. a meaningful English sentence. The thought is, to put it differently, one by which the institutional fact of language is (partly) created. The fact that language is an essential component of the existence of that particular institution which is the institution of language should not be very surprising. What remains to be shown, however, is that language is an essential component of other institutions such as money, kingdoms and cocktail parties.

The third argument I will refer to as the "there is nothing else there but linguistic symbols" argument. The argument consists in an analogy between the scoring of points in games and institutional reality. Searle explains that "a touchdown counts six points" is not a thought that one could have without linguistic symbols. Why not? Because, according to Searle, "points can only exist relative to a linguistic system for representing and counting points" (CSR, p. 66). Similarly, the analogy goes, "the yellow line counts as a frontier" is a thought that one could not have without linguistic symbols and the reason is that a frontier can only exist relative to a linguistic system for representing.

There are at least two ways of resisting the argument. One is to question the adequacy of the analogy, that is, on the possibility of treating frontiers, Kings, and money like touchdowns. Although games involve status functions and constitutive rules, they are not full-blown institutions, and so we should consider any analogies between the two with caution. I will examine this line of thought in the last section of this paper. The other line of reply, which is the one I examine now, is to refute the view according to which games could not be played if agents did not have linguistic ways of representing the scores. It is true that the game could not be played if points could not be counted and it is also a noticeable fact that numbers are in this linguistic system the way to represent and count these points. But the question is whether there is a non-linguistic way of representing and hence of counting those points? What if I fold up one of my fingers every time a point is made and use that registration device? Wouldn't it be an alternative way of representing and counting the points? While it surely would be less reliable (I need to keep the fingers fold up until the end of the game) and more limited (I only have 20 fingers) than using numbers (by either writing them down or by uttering them), yet it is, *Pace* Searle, a non-linguistic way of scoring points[446].

But Searle further claims that "if you take away all the symbolic devices for representing points, there is nothing else there" (CSR, p. 66). Now there are various ways of interpreting what Searle means by "there is nothing else there". Maybe he means that no point could be scored if there was no way of representing them at all,

[446] Searle does recognize the possibility of counting points by using "some other symbolic devices other than actual words" and gives as an example the possibility of "assembling piles of stones, one stone for each point." But Searle astoundingly adds that in this case "the stones would be as much linguistic symbols as would any others". But to my knowledge, stones are not words and although we may imagine a new language within which words would be stones of various shapes, such a language remains to be invented and until it is so, stones are not words.

whether linguistic or non-linguistic. The claim is however dubious. Take away all the representative devices available and, still, it remains the case that one team's score is raised by 6 points if one of its players makes a touchdown. The lack of representative devices does not change anything regarding that fact.

Searle will maybe reply that, being short of a symbolizing device, the players (or some authority attending the football game) need at least to represent to themselves the added points for the latter to exist. After all, it remains an uncontested claim that no point could ever be scored if there was no one to think that there are scored points. But even this apparently credible claim can be challenged. Suppose that a football game takes place and that one of the teams makes a touchdown. According to the rule of the game, the team now has six more points but suppose furthermore that precisely when the touchdown is made, no one registers it either by changing the numbers of the scoring board nor by folding his fingers or by using any other devices. To explain such failure, we can imagine that everyone is suddenly struck by a short period of amnesia (an effect of having taken enhancing drugs, for example) and that as a consequence no one is able to represent the new score to herself. Would it imply that no points have been made at all? Would it imply, as Searle says, that "there is nothing else there" that happened? My intuition is that it would not have such radical consequence. In spite of the collective amnesia, one of the team did score six more points and it is unfair that no one has counted them.

Searle might here be willing to invoke the type-token distinction in order to restrict the language dependency to the type. Points as token can be scored without anyone representing them as having being scored. But what is true of points as tokens is not true of points as types which not only need to be represented to exist, they also require words or other markers to exist. Here is how the argument can be addressed. In order to justify his claim, Searle rightly notes that the scoring of points is not something to be seen in addition to a man crossing the line carrying a ball. He also correctly observed that "points are not 'out there' in the way that planets, men, balls and lines are out there" (CSR, p. 68). But it does not follow that, in contrast to planets, points are nothing but words. It does not follow that "points are not something that can be thought of or can exist independently of words or other sorts of markers" (CSR, p. 68). For even if we didn't have words or, for that matter, any other sorts of symbols to refer to points, still the latter could be scored as long as they are represented (and mutually known as such) in everyone's mind.

At this point, McGinn's skepticism about the need of language remains vindicated. Let us see whether such skepticism also resists the two other defenses that Searle provides in favor of the dependency of institutions on language.

2. Status function indicators

As Searle repeatedly notices, the switch from the X to the Y is not visible. This is because the object that is referred to as a Y is not physically different from the object that is referred to as an X. Searle says that "the existence of institutional facts cannot in general be read off from brute physical facts of the situation" (CSR, p. 119). As he also says "there is nothing in the physics of the situation that makes [the fact that the man holding the ball has scored a touchdown] *apparent*" (CSR, p. 72, my emphasis).

However invisible it may be, the move from X to Y does take place and what makes it possible is our capacity to represent the X as a Y. Representation thus constitutes a crucial step in the existence of status function. "The only way to get to the Y status function", Searle says, "is to represent the X object as having that status" (MLS, p. 154). It is important to observe that the representation involved in the move from X to Y is a representation of something *as* something else. The move requires the capacity to represent the object designated by the X term *as* a Y. Note that the object referred to as X is not physically different from the object referred to as Y. Hence the invisibility of the move from the former to the latter.

There is an important feature of representations that Searle does not stress. It is the fact that representations can either be public or private. The move from X to Y can either be something that is accessible to everyone or something that we represent to ourselves. A case where the representation can remain private is, for example, the case where John Searle imposes on the full moon the function of indicating to him that it is time he trimmed his sideburns[447]. Obviously, Searle's private representation of the full moon as an aide-mémoire is all that is needed. He needs not inform other Californians of the status function he personally imposes on the full moon. A similar case is the use of a piece of string tied around one's wrist as a reminder. The string will serve as a reminder if and only if someone represents it to herself as a reminder. But that person need not share her representation with anyone. She need not inform anyone of the special use she makes of that piece of string[448].

Using the full moon, a piece of string, or anything else as a reminder is however not an institutional fact. These are illustrations of solitary acts. In the case of institutional facts, Searle could reply, the move from X to Y exists in so far as it is publicly available to everyone and thus *publicly* represented as existing. The reason why the new status needs "markers" is, according to Searle, "because, empirically speaking, there isn't anything else there"[449] or, as he says, because "there is no way to read off the status function Y just from the physics of the X". According to Searle, the invisibility of all constitutive rules explains the need for markers. If status functions were physically visible, just looking at the Y object would be sufficient.

The argument however does not withstand scrutiny. First of all, it is possible to dispute the absence of any visible difference between the object as an X and the same object as a Y. Searle hastily assumes that any visible difference would have to be a physical difference between the objects as a X and the objects as a Y. But it need not be so. What makes the object as an X visually different from the same object as a Y could alternatively be the sort of things we respectively do with X and with Y. There is no physical difference between cigarettes serving the function of providing relief from a nicotine craving, on the one hand, and cigarettes used as a medium of exchange, on the other hand. In both cases, cigarettes are physically identical. It does not follow that the following of the constitutive rule of counting cigarettes as a medium of exchange is itself entirely invisible. This is because the physical features of cigarettes are not all what there is to see when cigarettes are used as a medium of exchange. Agents behave differently when they smoke cigarettes and

[447] McGinn (1995) explores this example in his review of *CSR*.

[448] There is an apparent paradox here. A reminder is a helpful device for anyone who needs to be reminded of things. Yet there is one thing that one needs to remember when using a piece of string as a reminder which is the special use one makes of the string. That special use cannot be read off the physical features of the string. I suppose that its unusual location, that is, around the wrist, here helps its user to remember the special use she makes of it.

[449] CRS, p. 69.

when they use them as money and these behavioral differences are visible. These behavioral differences should not be surprising from Searle's perspective. As he stresses himself sometimes, constitutive rules modify the range of things that agents can *do*. The constitutive rule of the money system allows agents to buy things with cigarettes, something they could not do before. The fact that agents do not behave similarly when they use cigarettes as a nicotine provider as when they use them as a medium of exchange may indicate something to any newcomer. To be sure, these behavioral differences may be insufficiently reliable as a representative device to provide knowledge of the constitutive rule that creates the money system. It might not be easy to infer that agents count cigarettes as a medium of exchange from the fact that agents purchase them although they do not smoke them. The point is that, besides the physical similarities between X and Y, there are behavioral dissimilarities between the two situations, and these behavioral dissimilarities do not make the X object and the Y object entirely equivalent from an "empirical" point of view.

Status indicators are means by which "we impose intentionality on entities that are not intrinsically intentional" (CRS, p. 99). A status indicator is a representative device that allows an entity to represent something beyond its physical features.

But the function of status indicator is itself a status function. Uniforms, crowns, and wedding rings do not play their function, that of indicating that certain human beings have such and such status, in virtue of their physical features. Crowns are not intrinsically intentional. So we first have to impose the power to represent on crowns for crowns to be able to perform their function of representation.

Be that as it may, Searle would perhaps reply that there is a need in the special case of institutional fact, for a public way of representing the various Xs as Ys. He will recall that kings, universities and cocktail parties are not like pense-bêtes in that they involve a group of individuals who all need to be aware of the status functions. Echoing Reid's distinction between solitary acts and social acts, Searle argues that status indicators are the means by which everyone is informed of these status functions. Let us here quote Reid:

> A man may see, and hear, and remember, and judge, and reason: he may deliberate and form purpose and execute them, without the intervention of any other intelligent being. They are solitary acts. But when he asks a question for information, when he testifies a fact, when he gives a command to his servant, when he makes a promise, or enters into a contract, these are social acts of mind, and can have no existence without the intervention of some other intelligent being, who acts a part in them. Between the operations of the mind, which, for want of a more proper name, I have called *solitary*, and those I have called *social,* there is this very remarkable distinction, that, in the solitary, the expression of them by words, or any other sensible signs, is accidental. They may exist, and be complete, without being expressed, without being known to any other person. But, in the social operations, the expression is essential. They cannot exist without being expressed by words or signs, and known to the other party (Reid, 1969, 437-438[450]).

Suppose you and I are playing chess and a pawn is missing. "Let us count that coin as a pawn", I suggest to you as a way of replacing the missing pawn. The utterance is the representative device, audibly specifying to the other party the fact that the coin now counts as a pawn. Is the need to be known by some other party what explains the difference between the cases (e.g. all institutional facts) where a representative device is needed and cases (e.g. pense-bête) where it is not? No, it is

[450] Quoted by Mulligan (1987).

not. As I intend to show the difference between these two cases does not map onto the difference between institutional acts and solitary acts. The reason rather is that sometimes various brute facts qualify as the sort of X that is to be counted as a Y (as a reminder, as a pawn, as a king, etc.), increasing the risk of mistakes. Sometimes the wrong X may be mistaken as a Y. Making the move visible thus seems to be an efficient way of avoiding these confusions. Suppose, for example, that we are invited to Versailles and the king enters the room in which all the guests are gathered. If everyone can correctly represent to themselves a particular person as the king, he need not be marked out as such. For doing so would add nothing to what is already quite familiar to everybody. Making the move visible would be redundant. Let us however imagine that our knowledge of what the king looks like is based on the official portraits of the king. The latter are however too flattering to be reliably informative. Since confusion is in this case possible, the declaration: "Here comes the king!" that accompanies his entrance is more than helpful. Linguistic markers are obviously good ways of avoiding the possibility of taking the *wrong* X as a Y.

Let us investigate the example further and ask where the possibility of confusion — of taking the wrong X as a Y — comes from? One explanation is that not everybody was involved in the initial decision to take *that* X as a Y. There are those who made the choice and those who only (either reluctantly or enthusiastically) accept it and the latter obviously need to be notified of the choice of the former. The chess example also fits this explanation. Suppose that we never find the missing pawn and leave the coin in the chess box as a substitute for this missing pawn for future game sessions. Now a representative device, i.e. something visibly specifying that the sharpener is to be counted as a pawn, would be more than useful in case someone who was not part of the initial decision intended to use that chess game. So if the move from X to Y needs to be made visible, it needs to be so exclusively for mere acceptants, and not for the more august legislators, in order to count as such a move. This way the acceptants are made aware of the constitutive rule that has been determined and can represent to themselves the right X as the Y. The necessity of markers is, on such hypothesis, a consequence of the division within the society between those who makes the constitutive rule and those who merely accept it.

Interestingly, the explanation in terms of avoiding confusion does not pertain exclusively to institutional facts. Confusion can very well threaten the private imposition of status function. Suppose that I have a few screwdrivers at home and that I want to use one of them as a reminder. Unless I mark *the* screwdriver on which I impose that status function — unless I managed to make it visibly different from the others — there is no way that this screwdriver will be able to perform its reminding function. So what do I do? I hook it on the door. The hooking, I believe, is my way of marking the screwdriver-as a-reminder to make it different from the screwdrivers-as-screw-drivers and to avoid any confusion. So the hooking is a status indicator.

Suppose, as another example, that some trees are attributed the function of signaling that all drivers must slow down. Unless the drivers agree about which trees, among those standing along the road, have such status function, the rule cannot be followed. But agreeing is not enough, we additionally need to mark the trees to which a signaling function is assigned. This is because unless we have a means of distinguishing the regular trees from the signaling trees (by painting the latter in red for example, or, in a more green-friendly fashion, by deciding that only a certain type of trees, palm-trees, for example, will play the signaling function), the rule cannot be properly followed. To sum up, status indicators are needed when confusion is

possible and the latter arises when there are many Xs with which the qualified X that must be counted as a Y can be confused.

There is a way of verifying the adequacy of the proposed explanation. If it is correct, it would have the following implication: when the X in the formula "X counts as Y in C", is exemplified by one token only, the need for status indicators should just vanish. Does the full moon example not precisely show this? When Searle assigns to it the function of reminding him to trim his eyebrows, does he need to mark the full moon with a status indicator? No. The reason is that there is no full moon besides the one to which he assigns the function of reminding him to trim his eyebrows that might confound him about the time he must proceed with the trimming. The fact that the full moon is uniquely instantiated — the fact there is only one full moon a month — explains why Searle does not need to find a status indicator for marking the full moon as a reminder. Again, status indicators are needed because, sometimes, there are many Xs to which a constitutive rule could apply whereas we want to restrict the application of the constitutive rule to some of these X's only.

In sum, the need to find a status indicator shows up when the formula "X counts as Y" does not apply to all token of Xs. Not all human beings are kings, judges, or policemen. Unless we know which Xs, among all the Xs, count as Y in C, one must mark those that do have the function with a status indicator. On the proposed account, the need for status indicators is unrelated to the fact that "there is no way to read off the status function Y just from the physics of the X". It is not because judges, policemen and kings have correlatively no existence apart from our representation that we need to find a representative device. We need to find a representative device because judges, policemen and kings are, physically speaking, human beings and that not all human beings are judges, policemen and Kings. The only way to make a difference between human beings that are not judges and human beings that are judges is to mark the latter with a status-indicator (e.g. their uniform) that eliminates all possible confusion. Nor does the need to publicly mark institutional status functions have anything to do with the fact that, unlike private status functions, institutional facts involve more than one person. Objects that are privately used as reminders need to be similarly marked in order to avoid confusion.

Now the declaration: "The King is there!" [451] is certainly one efficient way to make the move from Louis Dieudonné to Louis XIV visible. Just as the words "shell money" indicate in a publicly available way that a status function, that is, serving as a medium of exchange, is attached to cowry shells. Among the various features of language, the capacity to symbolize plays a crucial role in the existence of institutional facts. Words are like labels attached to physical entities. They signal to everyone that a certain physical entity, an X, is in fact a Y, that is, an institutional entity. They are tags warning us of the special function that an entity plays, a function unrelated to its physical features. Searle calls these tags, "status indicators", "markers" or, more prosaically, "representative devices".

Words, however, are only metaphorically tags on institutional facts. For words inform us about institutional facts in the form of uttered declarations. Unlike tags, which are meant to be *seen*, notice that declarations do not add visibility to institutional facts. They make them more perceptible by making them distinctively *audible*. They add sounds, rather than visual cues, to institutional facts.

[451] This may not be the right formulation of the declaration.

To represent the X as having a status function is what status indicators do but status indicators come in various types. There are many ways, besides using Y terms, to make the move from X to Y noticeable. Having the king wear a crown is, for example, one alternative. Uniforms, crowns, and wedding rings are symbolic ways of marking the difference between human beings, on the one hand, and kings, policemen, judges and spouses (CRS, p. 120), on the other hand. Uniforms, wedding rings and words like "money", "king", "pawn" are like tags attached to certain material objects, indicating to everyone that these material objects play a function, one that cannot be read off from their physical features.

However, the view that language is constitutive of institutional facts entails that only *linguistic* markers are essential to the latter. Something could not be money unless it is *verbally* referred to as "money". Unless Searle tells us why linguistic symbols are exclusively powerful in making the move from X to Y, visible, his second argument is, at best, incomplete.

3. Status Function Declarations

We have so far dealt with the representations involved in all constitutive rules as if they were ordinary mental states. We have assumed that these representations were not different from, say, beliefs and that, as a consequence, their propositional contents were mere expressions of those mental states. On such an approach, the proposition "cowry shells count as money" is true in virtue of its ability to reflect the belief that each Ojibwas holds in this regard. The representation has, on this approach, a word to world direction of fit, as Searle would phrase it, to the extent that its truth depends on the accuracy with which these representations describes our beliefs. Its utterance consequently appears as something that is added to the representation as an optional extra.

This is however a mistaken way of construing them. The representations (the thoughts, the beliefs, the opinions, etc.) involved in a constitutive rule have an additional special power in consideration of which their classification as simple mental states sounds inaccurate or, at least, only in part true. Unlike the belief that the grass is green or the thought that I am in pain, the representations of an X as a Y have the power of creating the very reality that they describe. The content of the constitutive rule by means of which we, say, impose on cowry shells the function of media of exchange does not only represent our inner thought. It also changes the reality to the extent that there is now a new class of facts — selling, buying, storing, etc. that now can take place. Because the representations do not only reflect what agents think but also aim to "change the world by declaring that a state of affairs exists and thus bringing that state of affairs into existence" (MSW, p. 12), their word-to world direction of fit combines with a world-to-word direction of fit.

Recognizing such a power in the representations involved in the move from X to Y this way is to afford them a place in the familiar class of speech acts. So the claim about the dependency of institutional facts on language turns out to be a claim about the dependency of institutional facts on speech acts. And because of their double direction of fit, the representations involved in the move from X to Y nicely illustrate a subclass of speech acts, namely declarations. To be declared married is essential to being married. Just as to be declared a leader is essential to being a leader and to be declared a university is essential to being one. On Searle's terms, the

declarations that are at stake in the creation and maintenance of money, of banks and of universities are "Status Function Declarations" (hereafter SFD) and have the following structure: "We (or I) make it the case by Declaration that the Y status function exists." (CSR, p. 93).

Unlike mental states, the utterance is essential to speech acts. Unlike pain, which can occur independently from its expression in the form of an utterance (i.e. "I am in pain"), ascribing a status function is an experience that could not occur without its utterance. The reason is that, unlike the experience of pain, which does not need to be known, the SFD needs to be grasped[452]. There is nothing about declarations that could rightly be taken as the mere expression of a belief. Declarations essentially have a public dimension, in the form of their utterance, in addition to reflecting some private thoughts. SFDs are uttered, in other words, inasmuch as they necessarily have recipients. As Mulligan explains, "here the experience is not possible without the utterance. And the utterance for its part is not some optional thing which is added from without, but is in the service of the [speech act], and is necessary if the act is to carry out its function of making itself known to the other person"[453].

Yet if the point of the utterance is to make the act known to the other person, couldn't the latter be shouted at using something other than spoken words? In fact, they could. On Searle's view, "wearing a wedding ring or a uniform is performing a type of speech act" (*CSR,* p. 120). Such a liberal conception of speech act is also clearly the one he supports when he claims that "any intentional movement can be a speech act provided it is performed with certain sets of semantic intentions that are communicated to the hearer". "Speech act", he continues, "is a quasi-technical term that means, roughly, 'a meaningful linguistic act that is intended to communicate propositional content with a certain force from speaker to hearer, which may be spoken, written, or conveyed in some other symbolic form'"[454]. So a mere gesture can very well be the sort of act by which a speech act is performed as long as it is intended to communicate the existence of a status function.

But then it seems that what matters for being a speech act is the intention that accompanies the act and not the various expressions by which the act is externally reflected. The same gesture may, in some circumstances, be the expression of an act that is not a speech act. I can either push a beer as a gesture of disgust or because I want it to be yours. Although both gestures are acts, only the latter is a speech act. Surprisingly enough, speech turns out to be not so essential to speech acts after all and this is because, besides saying it out loud, there are many silent ways of making a status function known. Such a finding should be very welcome. In light of it, Searle's view about the central role of language in the existence of institutional facts finally accommodates McGinn's skepticism well.

Searle's third argument in favor of the central role of language focuses on the change that the move from X to Y brings about in reality. No real change could ever take place if the move from X to Y were only a matter of an inner representation. The change, it can be argued, would have been as non-existent as the one involved in the perceptual shift involved in seeing the duck-rabbit picture as a duck or as a rabbit. But the move from X to Y is different to the extent that a new entity — a king, an auction, a money bill or a judge — is introduced into the world.

---

[452] Cf. Reinach [1913], 1989.
[453] Mulligan, 1987, p. 32.
[454] Searle (2011).

In its present form, however, Searle's third argument leaves two sorts of cases unresolved. In particular, it does not rule out two already encountered cases, namely reminders and football games, although they are intuitively not relevant examples of institutional facts. In the last two sections, I show how a refined conception of the sort of declarations on which institutions depend helps us deal with these two cases.

4. Undeclared status functions

Sometimes, the representation involved in the ascription of a status function can perfectly remain *un*declared in any of the linguistic or behavioral ways of understanding what declarations can be. These are the various cases of reminders discussed earlier. While both reminders and frontiers involved the imposition of status function, Searle's third argument (which is based on the observation that the reality has changed, i.e. a new entity has been introduced into it) does not tell us why only in the case of frontiers such ascription needs to be a matter of a declaration. Just as frontiers, reminders are newly created entity. Note also that it is not that the imposition of function is, in the case of reminders, silently declared. Rather such imposition takes place without any declaration at all, not even a tacit or implicit one.

In order to understand why reminders are different from frontiers, Reinach's pioneering theory of "social acts" (which anticipates in many ways Austin and Searle's speech act theory) turns out to be useful[455]. There is, in particular, a distinction Reinach makes between two types of acts, namely "self-directable" and "non-self-directable"[456] ones, in light of which the status function of reminders and frontiers can be set apart. Unlike all social acts, the act of using a reminder is a self-directable act inasmuch as the subject toward whom it is directed is identical with the subject of the act. Non-self-directable acts require by contrast an alien subject. We should now see why SFDs are illustrations of social acts as non-self-directable acts. One does not declare — it just does not make sense — that a yellow line has the function of serving as a frontier to oneself. Like requests, admonishments, questionings, informings, answerings, SFDs have an announcing function that requires an addressee who also grasp their content.

Even more crucial to the difference between reminders and money is the impersonal feature that Reinach observes in a sub-category of social acts such as waiving a claim, revoking a promise and enactments. Reinach observes that, although these three acts are non-self-directable (we do not waive a claim to oneself) in the sense that they must be grasped by others to be fully performed, they nonetheless would be badly described as other-directed acts. This is because, unlike other social acts such as promises, these acts do not refer to any particular person. "Enactments" Reinach says, "do not have this necessary relation to other persons, just as little as do acts like waiving or revoking. Although these acts are addressed to other persons in being performed, their substance (*Gehalt*) lacks any personal moment"[457]. Likewise, there seems to be an *impersonal* dimension in the way SFD are addressed that makes them different from the personal way reminders are used. Note first that both reminders and frontiers presuppose a person or a group. Reminders refer to a person who is reminded and frontiers contain a reference to a group which is prohibited

[455] Reinach [1913], 1989.
[456] Reinach, [1913], 1989, 170.
[457] Reinach [1913], 1989, 170.

from trespassing on some territory. But whereas something is a reminder for somebody, something is a frontier for everyone. The SFD that is at the core of the latter does not pertain to *you* and *me*, inasmuch as they are not addressed to any persons in particular. They apply to everyone without anyone being individually addressed by those declarations. The kind of speech acts that pertains to institutional facts has this impersonal feature[458]. In light of it, the reason why reminders are not, despite their status function, part of such institutional reality, now makes sense.

5. Speech acts as pure acts

Searle's third argument may also be found lacking, as it does not account for the difference between full-blown institutional facts and more dubious cases. On Searle's conception, indeed, the act of scoring a goal is as much an institutional fact as the act of counting a yellow line as a frontier, to the extent that in both case a SFD captures the move from X to Y. The scoring of a goal in a football game depends in particular on the following SDF: "We (or I) make it the case by Declaration that the crossing of the line by the ball counts as the scoring of a goal". Such declaration would be a silent one akin to the declaration that "this is yours" that is involved in the pushing of a beer.

Are football games genuine institutions? I personally would welcome any theory that is able to account for the difference between the scoring of a goal and a presidential election. It seems to me that there is a difference between these two facts and that the difference is not only related to the level of gravity with which one of the two (and, of course, I will not specify which one) ought to be treated.

The difference can be grasped, I will now show, once speech acts are construed as illustrations of Moya's notion of "pure acts"[459]. Recall that pure actions are, according to Moya, actions that cannot be non-intentionally performed and that cases in point are greeting, signaling for a turn or marrying. Remember also that Moya explains the fact that pure actions are necessarily intentional in terms of the impossibility to distinguish them from their happening.

Consider now SFDs in light of the distinction between impure and pure acts. We should now see why SFDs belong to the latter category. Take, as an example, the following SFD: "I pronounce you married". It is impossible to sort out the happening — e.g. Catherine and Jules's marriage is pronounced — from the act itself — e.g. someone pronounces Catherine and Jules as married. Consider now the scoring of a goal. The possibility of separating the happening — a scored goal — from the act itself — the scoring of a goal is available. The possibility of distinguishing between the two comes from the possibility that the happening is mistakenly brought about. Mistakes are, by contrast, not something that is possible in the circumstance of a SDF. Under what credible circumstances, could it be possible to find oneself mistakenly declaring that two persons are pronounced married? Because it is subject to mistakes, scoring a goal is for this reason an impure act, one

[458] Otto Bruun objected to me that a declaration of marriage does not have the impersonal dimension which, on my account, is characteristic of the ascription of *institutional* status functions. As a reply I first grant that when the official declares: "I pronounce you married", he does personally direct his declaration to the bridegrooms, but also observes that the official does not personally address his declaration to any of the other members of the group within which the marriage takes place.

[459] Moya, 1990. Cf. chapter I, section 5. (pp. 51-55) for a more complete presentation of pure acts.

that falls outside the class of speech acts. No SFD turns out to be involved in the scoring of a goal.

In sum, the imposition of a status function may or may not stem from a status function declaration. Only in the latter case can a full-blown institutional fact be created. The latter case obtains, I have shown, when both (i) the possibility of being mistaken and (ii) the possibility of separating the happening from the act is impossible. Two conditions which the scoring of a goal clearly does not meet, consistent with our intuition that it is not a genuine institutional fact.

Conclusion

To the question "Why is language needed for the creation and maintenance of institutional facts?" Searle provides three different answers. One, institutional thoughts are too complex, he first argues, to be held without language. Two, institutional facts would remain invisible, he alternatively argues, if they were not publicly represented by means of some linguistic symbols. Three, the changes that are brought about each time an institutional fact obtains could not take place if those thoughts were not sub-types of speech acts, namely declarations, the latter being characterized by an external, non-psychological, side in virtue of which uttering them is doing something.

I have additionally brought into play further features of the sort of declarations that are involved in institutional facts, the fact that they are impersonally addressed and cannot be mistakenly performed, in order to highlight cases where status functions are ascribed outside institutional facts.

REFERENCES

Anderson, E. 2000. Beyond Homo Economicus: New Developments in Theories of Social Norms. *Philosophy Public Affairs* 29 (2), 170-200.

Andreoni, J, 1989. Impure altruism and donations to public goods: a theory of "warm-glow" giving, Working paper, University of Wisconsin.

Anscombe, E. 1957. *Intentions.* Oxford.

Aydinonat, E. 2006. Is the Invisible Hand Un-Smithian? A Comment on Rothschild. *Economics Bulletin* 2 (2), 1-9.

Aydinonat, E. 2008. *The Invisible Hand in Economics: How Economists Explain Unintended Social Consequences.* INEM Advances in Economic Methodology, London, Routledge.

Barry, N. 1982. The Tradition of Spontaneous Order. *Literature of Liberty* 5, 1-58.

Brennan G. & Pettit, P. 1993. Hands Invisible and Intangible. *Synthese* 94, 191-225.

Cartwright, N. 1994. Mill and Menger - Ideal Elements and Stable Tendencies, Poznan Studies in the Philosophy of the Sciences and the Humanities, 38, Idealization VI: Idealization in Economics, edited by B. Hamminga & Neil B. De Marchi, Amsterdam/Atlanta, GA, Rodopi, pp. 171-188.

Chan, J. & Miller, D. 1991. Elster on Self-realization in Politics: A critical Note. *Ethics* 102 (1), 96-102.

Cowen, T. 1997. Do Economics Use Social Mechanisms to Explain? *Social Mechanisms. An Analytical Approach to Social Theory,* edited by P. Hedström, R. Swedberg, Cambridge: Cambridge University Press, 125-146.

Craig, E. J. 1990. *Knowledge and the State of Nature.* Oxford, Clarendon Press.

Davidson, D. 1980. *Essays on Actions and Events.* Clarendon Press, Oxford.

Elster, J. 1978. *Logic and Society.* Chichester: Wiley.

Elster, J. 1983a. *Sour grapes*. Cambridge, Cambridge University Press.

Elster. J. 1983b. *Explaining Technical Change*. Cambridge, Cambridge University Press.

Elster, J. 1985a. *Making Sense of Marx*. Cambridge, Cambridge University Press.

Elster, J. 1985b. The nature and scope of rational-choice explanation. *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, edited by E. Lepore and B. Maclaughlin, Oxford: Blackwell, 60-72.

Elster, J. 1988. Economic order and social norms. *Journal of Institutional and Theoretical Economics* 144, 357-66.

Elster, J. 1989. Social norms and economic theory. *Journal of Economic Perspectives* 3, 99-118.

Elster, J. 1993. Why things don't happen as planned. *The Necessity of Friction,* edited by N. □ kerman, Heidelberg, Physica Verlag, 248-56.

Elster, J. 2007. *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*. Cambridge, Cambridge University Press.

Elster, J. 2009. *Le Désintérêt. Traité critique de l'homme économique,* Paris, Seuil, 2009.

Emerson M. O., Chai K. J., Yancey, G. 2001. Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans. *American Sociological Review,* Vol. 66 (6), 922-935.

Fehr, E. & Gachter, S. 2000. Cooperation and Punishment in Public Goods Experiments, *American Economic Review, American Economic Association* 90(4), 980-994.

Fehr, E. 2002. Why Social Preferences Matter. The Impact of Non-selfish Motives on Competition, Cooperation and Incentives. *Economic Journal* 112, C1-C33.

Feinberg, J. [1958] 2008. Psychological Egoism. *Reason & Responsibility: Readings in Some Basic Problems of Philosophy*, edited by Joel Feinberg and Russ Shafer-Landau, California, Thomson Wadsworth, 520-532.

Ferguson, A. [1767] 1782. *An Essay on the History of Civil Society*. 5th ed., London, T. Cadell.

Fogel, R., Enferman S. 1974. *Time on the Cross: The Economics of American Negro Slavery*. Boston, Little, Brown and Company.

Frey, B. 1997. *Not Just For the Money. An Economic Theory of Personal Motivation*. Edward Elgar Publishing Limited, Cheltenham.

Friedman, J. 2006. Comment on Searle's social ontology. The reality of the imaginary and the cunning of the non-intentional. *Anthropological Theory* 6 (1), 70-80.

Gambetta, D. 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge, Massachusetts, Harvard University Press.

Gaus, G., (forthcoming) Explanation, Justification, and Emergent Properties: An Essay on Nozickean Metatheory. *The Cambridge Companion to Nozick's 'Anarchy, State, and Utopia'*, edited by Ralf M. Bader and John Meadowcroft. Cambridge, Cambridge University Press.

Gilbert, M. 1989. *On Social facts*. Princeton, New Jersey, Princeton University Press.

Craig, E. J. 1990. *Knowledge and the State of Nature*. Oxford, Clarendon Press.

Grimen, H. 1994. Causally inefficient knowledge and functional explanation. *Social Science Information* 33, 117-127.

Güth, W. Schmittberger, R. Schwarze V., 1982. An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behaviour and Organization* 3 (4), 367–388.

Haller, M. 2000. Carl Menger's Theory of Invisible-hand Explanations. *Social Science Information* 39, 529-565.

Haller, M. 2002. Expliquer l'existence des institutions par la main invisible: Menger et après. *La philosophie autrichienne de Bolzano à Musil,* edited by Jean-Pierre Cometti et Kevin Mulligan, Paris, Vrin.

Haller, M. 2004. Mixing Economics and Ethics: Carl Menger vs Gustav Von Schmoller. *Social Science Information* 43(1), 5-33.

Hamowy, R. 1987. *The Scottish Enlightenment and the Theory of Spontaneous Order*. Southern, Illinois University Press.

Hayek, F. A. 1973. *Law, Legislation and Liberty: Rules and Order*. Vol. 1. Chicago: University of Chicago Press.

Hayek, F. A. 1948. *Individualism and Economic Order*. Chicago, University of Chicago Press.

Hayek, F. A. 1966. Lecture on a Master Mind: Dr. Bernard Mandeville. *Proceedings of the British Academy* 52, 125-141.

Hegel, F. 1975. *Philosophie de l'histoire,* translated by Jacques D'Hondt, Paris, Presses Universaires de France.

Henrich, J. Boyd, R. Bowles, S. Camerer, C. Fehr E. and Gintis H., 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford, Oxford University Press.

Hill, L. 1998. The Invisible Hand of Adam Ferguson, *The European Legacy,* 3 (6), 42-65.

Hindriks, F. A. 2003. The new role of the constitutive rule. *American Journal of Economics and Sociology* 62 (1), 185-208.

Hindriks, F. A. 2005. *Rules & Institutions. Essays on Meaning, Speech Acts and Social Ontology*. Alblasserdam, Haveka BV.

Hindriks, F. A. (forthcoming). Restructuring Searle's *Making of the Social World*. the *Philosophy of Social Science*.

Hirschman A. O. 1977. *The Passions and the Interests: Political Arguments For Capitalism Before Its Triumph*. Princeton, New Jersey, Princeton University Press.

Hirshman, A. O. 1984. Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse. *American Economic Review* 72 (2), 89-96.

Hirschman 1991. *The Rhetoric of Reaction: Perversity, Futility, Jeopardy*. Cambridge, Massachusetts, The Belknap Press of Harvard University Press.

Holmes, S. 2009. Saved by Danger/Destroyed by Success. The Argument of Tocqueville's Souvenirs. *European Journal of Sociology* 50, 171-199.

Hull, D. 1988. *Science as a Process.* Chicago, University of Chicago Press.

Hull, D. 1997. What's Wrong with Invisible-Hand Explanations? *Philosophy of Science,* 64 (Proceeding), S117-S126.

Hume, D. [1740], 2000. *A Treatise of Human Nature*, edited by David Fate Norton and Mary J. Norton, Oxford/New York, Oxford University Press.

Iwai, K. 2001. Evolution of Money. Evolution of Economic Diversity edited by U. Pagano and A. Nicita, London, Routledge, 2001, 396-431.

Kaufmann, L. 2005. Self-in-a-Vat. On John Searle's Ontology of Reasons for Acting. *Philosophy of the Social Sciences* 35 (4), 447-479.

Keller, R. 1994. *On Language Change: The Invisible Hand in Language.* London and New York, Routledge.

Lagerspetz, E. 1995. *The opposite mirrors: An essay on the conventionalist theory of institutions.* Dordrecht, Kluwer Academic.

Lewis, D. 1969, *Convention.* Cambridge, Harvard University Press.

Little, Daniel, 2011 (February 27th). Searle on Social Ontology (Web log post). Retrieved from http://understandingsociety.blogspot.com/2011/02/searle-on-social ontology.html

Lukes, S. 2006. Searle and his critics. *Anthropological Theory* 6 (1), 5-11.

Mäki, U. 1990a. Practical Syllogism, Entrepreneurship and the Invisible Hand. A critique of the analytic hermeneutics of G. H. von Wright. *Economics and hermeneutic,* edited by Don Lavoie*,* London, Routledge.

Mäki, U. 1990b. Scientific realism and Austrian explanation. *Review of Political Economy* (2), 310-344.

Mäki, U. 1990c. Mengerian Economics in Realist Perspective. *History of Political Economy,* Annual Supplement to Vol. 22, 289-310.

Mäki, U. 1997. Universals and the Methodenstreit: A reexamination of Carl Menger's conception of economics as an exact science. *Studies in History and Philosophy of Science,* 28, 475-495.

Mäki, U. 2001. Economic ontology: What? Why? How?, *The Economic World View. Studies in the Ontology of Economics,* edited by U. Mäki. Cambridge, Cambridge University Press, 3-14.

Mandeville, B. [1723] 1988. *The Fable of the Bees or Private Vices, Publick Benefits,* 2 vols. with a critical, historical, and explanatory commentary by F.B. Kaye, Indianapolis, Liberty Fund.

Marty, A. 1908. *Untersuchungen zur Grundlegung der allgemeinen Grammatik und Sprachphilosophie.* Halle a.S., M. Niemeyer.

Mathiessen K. 2002. Searle, Collective intentions, and individualism. *Social Facts & Collective Intentionality,* edited by Georg Meggle, 185-204.

Menger, C. 1892a. On the origins of money. Translated by C. A. Foley. *Economic Journal* 2, 239-255.

Menger, C. 1892b. La monnaie mesure de valeur, *Revue d'économie politique* 6, 159-175, reproduced in Gilles Campagnolo, *Carl Menger entre Aristote et Hayek, Aux sources de l'économie moderne,* Cnrs Editions, Paris, 2008, 206-220.

Menger, C. [1883], 1985. *Investigations Into The Method of the Social Sciences with Special Reference to Economics.* Formerly published under the title: *Problems of Economics and Sociology,* With a new introduction by Lawrence H. White, Edited by Louis Schneider, Translated By Francis J. Nock, New York and London, New York University Press.

Menger, C. [1871], 2007. *Principles of Economics.* Translated by J. Dingwall and B. F. Hoselitz, foreword by Peter G. Klein, with an introduction by Friedrich A. Hayek, Ludwig von Mises Institute, Auburn, Alabama.

McGinn, C. 1995. Contract with Reality, *The New Republic*, May 22.

McGinn, C. 2010. Is Just Thinking Enough? *New York Review of Books*, November 11th.

McGinn, C. 2011. Letter to the Editor, Colin McGinn replies. *New York Review of Books,* February 24th.

Millar, J. [1771], 2006. *The Origin of the Distinction of Ranks; or, An Inquiry into the Circumstances which give rise to Influence and Authority in the Different Members of Society,* edited and with an introduction by Aaron Garrett, Indianapolis, Liberty Fund.

Miller, S. 2005. Artefacts and collective intentionality. *Techne, Journal of the Society for Philosophy and Technology* 9 (2), 52-67.

Moural, J. 2008. "Language and Institutions in Searle's *The Construction of Social Reality",* in *The Mystery of Capital and the Construction of Social Reality,* edited by Barry Smith, David M. Mark, and Isaac Ehrlich, Chicago: Open Court, 97-112.

Moya, C. 1990. *The Philosophy of Action.* Cambridge, Polity Press.

Mulligan, K. 1987. "Promisings and other Social Acts: Their Constituents and Structure » In, ed. K. Mulligan, *Speech Act and Sachverhalt: Reinach and the Foundations of Realist Phenomenology,* Dordrecht, Nijhoff, 29-90.

Nadeau, R. 2003. Carl Menger et le conflit des méthodes [unpublished manuscript].

Nadeau, R. 2005. Carl Menger et la méthodologie de l'économie politique, *Economies et Sociétés, Histoire de la pensée économique,* vol. 36, no. 6, 1187-1218.

Nietzsche [1887] (1994). *On the Genealogy of Morality.* Translated by Carol Diethe, Cambridge, Cambridge University Press.

Nozick, R. 1974. *Anarchy, State and Utopia,* Oxford, Blackwell, Basic Books.

Nozick, R. 1994. Invisible-Hand Explanations*, The American Economic Review.*

Olson, M. & McGuire, M. 1996. The Economics of Autocracy and Majority Rule: The Invisible Hand and the Use of Force, *The Journal of Economic Literature*, 34(1), 72-96.

Petsoulas, C. 2001. *Hayek's Liberalism and its Origins: His Idea of Spontaneous Order and the Scottish Enlighenment.* London, Routledge Studies in Social and Political Thought.

Postema. G. 2008. Salience Reasoning. *Topoi,* vol. 27, 41-55.

Pettit, P. 1990. Virtus Normativa: Rational Choice Perspectives. *Ethics* 100, 725-55.

Pettit, P. Brennan, G. 1993. Hands Invisible and Intangible *Synthese* 94, 191-225.

Pettit, P. 1993. *The Common Mind: An Essay on Psychology, Society and Politics*. Oxford University Press, New York.

Pettit, P. 1995. The Virtual Reality of Homo Economicus. *Monist* 78, 308-329.

Pettit, P. 1996. Functional Explanation and Virtual Selection. *British Journal for the Philosophy of Science* 47, 291-302.

Pettit, P. 1998. The Invisible Hand, *The Handbook of Economic Methodology* edited by J. B. Davis, D. Wade Hands, U. Mäki, Cheltenham, Northampton, 256-259.

Pettit, P. & Brennan, G. 2005. The Feasibility Issue. *The Oxford Handbook of Contemporary Philosophy*, edited by F. Jackson & M. Smith, Oxford University Press, 258-279.

Pettit, P. 2007. Resilience as the Explanandum of Social Theory. *Contingency* edited by Ian Shapiro and Sonu Bedi, NYU Press, New York.

Polanyi K. 1968. *Primitive, Archaic and Modern Economies*, edited by G. Dalton. New York, Doubleday.

Popper, P. 1962. *The Open Society and its Enemies*. New York: Harper Torchbooks.

Reid, T. (1969) [1788], *Essays on the Active Powers of the Human Mind,* Introduction by Baruch Brody, Cambridge, Mass., M.I.T.

Reinach, A. 1913. *The A Priori Foundations of Civil Law*. Translated by J. Crosby, *Aletheia* 3, 1–142.

Rosenberg, A. 1979. Can Economic Theory Explain Everything? *Philosophy of the Social Sciences* 9, 509-529.

Rosenberg, A. 1986. On the Explanatory Role of Existence Proofs. *Ethics* 97, 177-186.

Rosenberg, A. 2007. *Philosophy of Social Science*. Third Edition, Revised, Boulder, Westview/Harper Collins.

Rothschild, E. 2001. *Economic Sentiments, Adam Smith, Condorcet, and the Enlightenment*. Cambridge, London, Harvard University Press.

Rothschild, E. 1994. Adam Smith and the Invisible Hand. *American Economic Review,* 84, 319-322.

Ruben, D.-H. 1998. Philosophy of the Social Sciences, *Philosophy: A Guide Through the Subject*, edited by A. Grayling, Oxford University Press, Volume 2.

Schelling, T. C. 1960. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA, London, England.

Schelling, T. C. 1969. Models Of Segregation. *American Economic Review* 59, 488-493.

Schelling, T. C. 1971. Dynamic Models of Segregation. *Journal Of Mathematical Sociology* 1 (1), 143-86.

Schelling, T. C. 1978. *Micromotives and Microbehavior*. New York, London, W. W. Norton & Company.

Schelling, T. C. 1997. Models of Segregation. *American Economic Review* 59, 488-493.

Schmid, H.-B. 2005. Beyond Self-Goal Choice. Amartya Sen's Analysis of the Structure of Commitment and the Role of Shared Desires. *Economics and Philosophy* 21/1, 51-63.

Schmid, H.-B. 2008. Plural Action: Concepts and Problems. *Philosophy of the Social Sciences* 38(1), 25 54.

Schmid, H.-B. 2009. *Plural Action. Between Philosophy and the Social Sciences*. Springer Verlag, Contributions to Phenomenology (58).

Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. New York, Cambridge University Press.

Searle, J. 1990. Collective Intentions and Actions, *Intentions in Communication,* edited by P. Cohen, J. Morgan and M. Pollack, The MIT Press, Cambridge, Massachusetts, 401-415.

Searle, J. 1995. *The Construction of Social Reality*. New York, Free Press.

Searle, J. 1997. Responses to Critics of *The Construction of Social Reality*. *Philosophy and Phenomenological Research* 57 (2), 449-458.

Searle, J. 2001. Rationality in action. Cambridge Massachusetts, MIT Press.

Searle. J. 2005. What is an Institution? *Journal of Institutional Economics* 1 (1), 1-22.

Searle. J. 2006. Social ontology: Some basic principles. *Anthropological Theory* 6 (1), 12-29.

Searle. J. 2007. Social ontology: The problem and step toward a solution, *Intentional acts and institutional facts: Essay on John Searle's social ontology*, edited by S. L. Tshoatzidis, S. L. Dordrecht, Springer, 11-29.

Searle. J. 2008. Social ontology and political power. In *The mystery of capital and the construction of social reality*, edited by Barry Smith, David M. Mark, Isaac Ehrlich, Chicago, LaSalle, Illinois, Open Court, 19-34.

Searle, J. 2010. *Making the Social World: The Structure of Human Civilization,* New York, Oxford University Press.

Searle, J. 2011 (February 24th). In response to: "*Is Just Thinking Enough?*" from the November 11th, 2010 issue, Letter to the Editor in *New York Review of Books.*

Sen. A. 2002. *Rationality and Freedom*, Harvard, Harvard Belknap Press.

Shearmur. J. 2008. The construction of social reality: Searle, de Soto, and Disney, *The mystery of capital and the construction of social reality*, edited by B. Smith, D. Mark, I. Ehrlich, Chicago, Open Court, 53-78.

Simon, H. A. 1956. Rational choice and the structure of the environment. *Psychological Review* (63) 2, 129-138.

Smith A. [1759], 1982. *The Theory of Moral Sentiments,* edited by D.D. Raphael and A.L. Macfie, vol. I of the Glasgow Edition of the *Works and Correspondence of Adam Smith*, Indianapolis, Liberty Fund.

Smith A. [1795], 1982. *Essays on Philosophical Subjects,* edited by W. P. D. Wightman and J. C. Bryce, vol. III of the Glasgow Edition of the Works and Correspondence of Adam Smith. Indianapolis, Liberty Fund, 1982).

Smith, A. [1776], 1904. *An Inquiry into the Nature and Causes of the Wealth of Nations* by Adam Smith, edited with an Introduction, Notes, Marginal Summary and an Enlarged Index by Edwin Cannan, London, Methuen, 1904. 2 vols.

Smith, B., Wolfgang Grassl (eds.), 1986. *Austrian Economics: Historical and Philosophical Background*, New York, New York University Press, London/Sydney, Croom Helm.

Smith, B. 1995. *Austrian Philosophy: The Legacy of Franz Brentano,* La Salle and Chicago, Open Court.

Smith, B. 2003. John Searle: From Speech Acts to Social Reality, in B. Smith (ed.), *John Searle*, Cambridge, Cambridge University Press, 2003, 1-33.

Smith, B. 2010. Aristotle, Menger, Mises: An Essay in the Metaphysics of Economics, *History of Political Economy*, Annual Supplement to vol. 22 (1990), 263-288.

Sorensen, S. 2010. Veridical Idealizations (Unpublished manuscript).

Steiner, H. 1978. Can a Social Contract be Signed by an Invisible Hand? *Democracy, Consensus and Social Contract edited by* Birnbaum, P. *et all.,* London, UK & California, Sage Publications.

Stewart, D. 1858. *Biographical Memoirs of Adam Smith, L.L.D., of William Robertson, D. D. and of Thomas Reid, D. D. Read before the Royal Society of Edinburgh. Now collected into one volume, with some additional notes*, in *Collected Works*, vol. X, Edinburgh.

Sugden, R. 2000. Team Preferences, *Economics and Philosophy* 16 (2), 175-204.

Sugden, R. 2002. Credible worlds: the status of theoretical models in economics, *Fact and Fiction in Economics: Modeals, Realism and Social Construction* edited by U. Mäki, Cambridge University Press, 107-136.

Sugden, R., Zamarrón I. 2006. Finding the Key: The Riddle of Focal Points, *Journal of Economic Psychology* 27, 609-621.

Sugden, R., Gold, N. 2007. "Collective Intentions and Team Agency", *Journal of Philosophy* 104, 109-137.

Tieffenbach, E. 2002. La philosophie de l'histoire d'Edmund Burke (1729-1797): de l'ordre spontané à l'*anarchie méthodique, Annales Benjamin Constant,* n° 26, 176-206.

Tieffenbach, E. 2003. De la 'main invisible' à la 'Ruse de la Raison': traduction romantique d'une idée des Lumières, *Dénouement des Lumières et Invention Romantique* edited by G. Bardazzi*,* A. Grosrichard*,* Genève, Droz, 47-67.

Tieffenbach, E. 2010. Searle and Menger on Money, *Philosophy of the Social Sciences* 40 (2), 191-212.

Tieffenbach. E. 2011. The Sounds of Institutional Facts, *Philosophical papers dedicated to Kevin Mulligan,* edited by A. Reboul. Retrived from http://www.philosophie.ch/kevin/festschrift/

Tocqueville, A. de [1893] 1978. *Souvenirs,* Paris, Gallimard.

Tuomela, R. 1977. *Human Action and Its Explanation,* Dordrecht.

Tuomela, R. 1984. *A Theory of Social Action,* Dordrecht.

Tuomela, R. 2003. Collective acceptance: Social institutions and social reality, *American Journal of Economics and Sociology* 62 (1), 123-64.

Turner, S. P. 1995. Searle's social reality, Review essay on *The construction of social reality*. *History and Theory* 38, 211-31.

Ullmann-Margalit, E. 1977. *The Emergence of Norms,* Clarendon Press, Oxford University Press.

Ullmann-Margalit, E. 1978. Invisible-Hand Explanations, *Synthese,* 39, 263-291.

Ullmann-Margalit, E. 1997. The Invisible Hand and the Cunning of reason, *Social Research* 64 (2), 181-198.

Velleman, D. T. 1997. How To Share An Intention, *Philosophy and Phenomenological Research* 57 (1), 29-50.

Vernon, R. 1979. Unintended Consequences, *Political Theory* 7 (1), 35-73.

Viskovatoff, A. 2003. Searle, Rationality, and Social Reality, *American Journal of Economics and Sociology* 62 (1), 7-43.

Von Kempski, J. 1992. Zur Logik der Ordnungsbegriffe besonders in den Sozialwissenschaften, *Prinzipien der Wirklichkeit.* Schriften 3, Frankfurt am Main, Suhrkamp, 339-367.

Von Wright, G.H. 1963. *The Logic of Preference,* Edinburgh.

Weber, M. 1949. *Methodology of the Social Sciences,* translated and edited by E. A. Shils and H. A. Finch, Chicago, Free Press.

White, L. 1985. "Introduction" to *Carl Menger, Investigations into the Method of the Social Sciences with Special Reference to Economics*, New York, New York University Press, vii-xviii.

Williams, B. 2002. *Truth and Truthfulness.* Princeton, New Jersey, Princeton University Press.

Wray, B. W. 2000. Invisible Hands and the Success of Science, *Philosophy of Science* 67, 163-175.

Ylikoski, P., Mäkelä, P. 2002. We-attitudes and social institutions, *Social facts and intentionality*, edited by G. Meggle, Frankfurt, Dr. Hänsel-Hohenhausen AG.

Ylikoski, P. 1995. The Invisible Hand and Science, *Science Studies 8*, 32-43.

Zaibert, L. 2004. Toward meta-politics, *The Quaterly Journal of Austrian economics* 7 (4), 113-128.

Zelizer, V. 2005. *The Purchase of Intimacy*, Princeton, N.J., Princeton University Press.

Zùñiga, G. L. 2005. Truth in Economic Subjectivism. *Philosophers of capitalism: Menger, Mises, Rand, and beyond*, edited by Edward W. Younkins, Lanham, MD, Lexington Books, 133-141.