# Social Science Information

**Invisible-hand explanations: From blindness to lack of we-ness**

Emma Tieffenbach

The online version of this article can be found at:

Published by:

**$SAGE**

http://www.sagepublications.com

On behalf of:

**SSI**

Maison des Sciences de l'Homme

Additional services and information for *Social Science Information* can be found at:

**Email Alerts:** http://ssi.sagepub.com/cgi/alerts

**Subscriptions:** http://ssi.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

**Citations:** http://ssi.sagepub.com/content/52/3/450.refs.html

>> Version of Record - Aug 5, 2013

What is This?

# Invisible-hand explanations: From blindness to lack of we-ness

**Emma Tieffenbach**
University of Geneva, Switzerland

## Abstract

The unintendedness of the phenomenon that is to be explained is a constraint visible in the various applications and clarifications of invisible-hand explanations. The article casts doubt on such a requirement and proposes a revised account. To have a role in an invisible-hand process, it is argued, agents may very well act with a view to contributing to the occurrence of the social outcome that is to be explained, provided they see what they do as an aggregation of their individual actions rather than as something they jointly perform.

## Keywords

collective intentionality, conventions, David Lewis, invisible-hand explanations, joint action, unintended consequences

## Résumé

La non-intentionnalité du phénomène à expliquer est une contrainte évidente dans les diverses explications par la main-invisible ainsi que dans les clarifications qui en ont été présentées. L'article jette un doute sur une telle exigence et en propose une représentation révisée. Pour jouer un rôle dans un processus de main-invisible, affirme l'auteur, les agents peuvent très bien agir en vue de contribuer à l'occurrence de l'effet qu'il faut expliquer, à partir du moment où ils voient ce qu'ils font comme la simple agrégation de leurs actions individuelles plutôt que comme quelque chose qu'ils accomplissent conjointement.

## Mots-clés

action conjointe, conséquences inattendues, conventions, David Lewis, explications par la main-invisible, intentionnalité collective

**Corresponding author:**
Emma Tieffenbach, University of Geneva – Centre Interfacultaire en Sciences affectives, 7 rue des Battoirs, Geneva 1205, Switzerland.
Email: emma.tieffenbach@unige.ch

While investigating the nature of collective intentions, John Searle (1990) offers the example of some businessmen who are familiar with Adam Smith's theory of the invisible hand. Having learned in business school that the best way for them to help humanity is by pursuing their selfish interest, they consequently each form an intention to this effect. Searle intends this example to show the following: although the businessmen pursue the same ultimate goal and are mutually aware of each other's interdependent intention, they can nevertheless not be adequately described as acting together. Helping humanity is something each businessman intends to do, albeit individually rather than jointly. A we-intention, the example purports to show, is not the sum of I-intentions, however intricate and goal-converging the latter might be. It is, as Searle also says, a primitive notion.

The example deserves attention for another reason as well. It construes the theory of the invisible hand as an anti-We-ness theory. Wealth could be redistributed just as well, the example indeed shows, 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?' The invisible hand is presented as a theory that considers the capacity to act jointly as a superfluous component of the way the social world works.

This article expands on the second idea, that of defining the invisible hand as an anti-We-ness theory, by making use of the anti-reductivist view about collective intentionality that Searle, among others, defends. In other words, I show that 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. I show that approaching the invisible hand in this manner entails a deep revision of its 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. Or, rather: although unintendedness is a sufficient feature of invisible-hand consequences, it turns out not to be a necessary one.

## The standard account and its problem

The invisible hand, as I first specify, has two facets. It is, on the one hand, well known as a normative theory, one that some invoke in order to defend the privatization of public goods, free entrepreneurship, and restricting state intervention in the economic, cultural or political realms. It is also, on the other hand, an explanatory theory, one that conveys a particular understanding of social outcomes, independently of the way they should be obtained. I am presently interested in the other face of the invisible hand, that is, in 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. Nor do I say anything about the relation between the two sides of the invisible hand, that is, whether endorsing it as an explanatory device commits you in any manner to accepting it as a political principle. I focus on invisible-hand explanations (IHE) and in particular on one of the constraints, namely the unintendedness of the outcome that is to be explained, that one allegedly needs to meet in order to provide an explanation of that type.

Instead of the well-known paradigm cases – Carl Menger's account of money (1892), Robert Nozick's state-of-nature story of about how the state could arise (1974) or Thomas Schelling's checkerboard city (1969) – the less canonical example of John Millar's explanation of the end of slavery (2006 [1771]) is our illustrative case. It is an account which Millar, a member of the Scottish Enlightenment and disciple of Adam Smith, provides in the final chapter of his *The Origin of the Distinction of Ranks* (2006 [1771]: 262–282). In the beginning, Millar explains, 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 further 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. Slave 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, in Millar's view, when hiring slaves became less costly than coercing them.[1]

Millar's explanation is an invisible-hand explanation insofar as it dispenses with a designing intelligence (Ullmann-Margalit, 1978). The end of slavery, Millar shows, need not be first represented in someone's mind as a blueprint, and then be put into effect according to some establishment program. Millar describes the end of slavery as the result of a process during which agents were not guided though their choices by the Church, or by some Great Men. His account traces back the responsibility for the extinction of slavery to the slave owners, who acted in the light of what their personal interest dictated, unaided in this task 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 joint 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 subvert the system. There are no slave owners mutually agreeing to end the practice of slavery after having discussed its various advantages and disadvantages. Absent from the account is any Lockean express agreement between the participants. The end of slavery is the result of each slave owner's similar but uncoordinated choice to resort to a new behavioural control system, one based on material reward rather than compulsion.

Another feature of Millar's explanation is that it involves agents who are strikingly unaware of what each of their actions will ultimately bring about. Neither the slave owners nor their slaves have the end of slavery in mind while acting. Of course neither of them are entirely blind since they have goals and are able to assess whether they achieve them or not. But the goal they individually pursue happens to be different from the large-scale social pattern that they produce, and the latter is the outcome that is to be explained.

This is, by all accounts, what invisible-hand explanations do. Invisible-hand explanations are consensually defined as accounting for some social outcomes in terms of the actions of many individuals who, as Nozick says, do 'not have that outcome in mind'

(1974: 14, 19) while acting. In her path-breaking article on invisible-hand explanations, Ullmann-Margalit also takes the same restrictive view. She says that agents cannot be aware, and hence foresee the overall effects, of their actions (1978: 271). According to Raimo Tuomela's conception 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' (1984: 451). Uskali Mäki (1990) also subscribes to this view when he defines an invisible-hand consequence as one that lies outside the agents' 'sphere of intendedness'. Unintendedness is therefore a constraint visible in the various applications and clarifications of invisible-hand explanations. The standard account (SA) 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.

This way of characterizing the invisible hand raises a difficulty, however. It puts forward a constraint that is not obviously met by paradigm cases. The very capacity of invisible-hand explanations to persuasively present their explanandum as an unintended consequence has been questioned. Christina Petsoulas (2001), for example, refutes Friedrich von Hayek's (1948) inclusion of the Scottish Enlightenments within the tradition of spontaneous order on this ground. Mandeville, Hume and Smith, she argues, assumed too many reflective and purposeful aptitudes in the development of institutional rules for their accounts of institutions to qualify as invisible-hand explanations. Hillel Steiner (1978) and Gerald Gaus (2011) address the same kind of criticism to Nozick's explanation of the minimal state when they claim that it is not obvious that the rise of a minimal state is not among the goals of the dominant protective agency. As Gaus notes:

> The actions and reasoning of the dominant protective agency in prohibiting unauthorized enforcement by independents is too close to aiming at [the minimal state] to constitute a satisfying invisible hand explanation of [the minimal 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 a minimal, Lockean, statehood. (Gaus, 2011: 124)

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-minimal state to a minimal state seems more deliberately brought about than it is supposed to. Nozick's explanation of the minimal state is therefore too close to a social-pact explanation of the same social phenomenon to be a compelling illustration of an invisible-hand explanation. Leo Zaibert (2004) raises a similar point when he argues that sometimes the relationship between what agents do individually and what they as a group bring about is so obvious that their lack of awareness appears as mere stupidity. Describing the outcome as an unintended consequence has the inconvenient implication that it restricts the application of invisible-hand-type explanations to only societies of fools.

At the core of these various criticisms is the idea that many invisible-hand explanations involve agents who are not sufficiently or, at least, not clearly 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, it is argued, is not wide enough to allow the working of an invisible hand. The disagreement between invisible-hand explainers and their critics, it may be summarized, results from a clash of intuitions as to how far from the agents' view the consequence must be situated to count as unintended rather than as intended.

There are various ways of dealing with these criticisms. One is to examine whether the intended vs unintended consequences dichotomy can be disambiguated, i.e. to what extent its sense can be sharpened so that it proves to be an adequate classificatory tool after all. We need to figure out what level of theoretical knowledge agents who are led by the invisible hand shall be (dis)allowed to have. Because there is no clear-cut boundary between ignorance and awareness, however, there won't be any straightforward answer to this question. Tyler 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?'(Cowen, 1997: 135). Answers to this question are likely to reveal a clash of intuitions given the grey zone that stands between blindness and clairvoyance. The sphere of intendedness has fuzzy boundaries and is for this reason unlikely to convincingly play the classificatory role it is supposed to.

Another line of inquiry is to examine whether the constraint is necessary and, in particular, whether it might not be too restrictive. Following the latter strategy, we ask to what extent the outcome can enter agents' 'sphere of intention' without ruling out the operation of an invisible hand. Having proceeded by submitting the standard account to various revisions of roughly increasing force, until a clearly inappropriate case for the invisible hand is encountered, I reveal that quite a few revisions need to be tested before such a boundary case is found.

## To have the overall outcome in mind

Let us consider the SA from this revisionary perspective and see what would happen if agents had the final outcome of their actions in mind while acting. Suppose, to take our illustrative starting example, that the slave owners can recognize that their decision to hire their slaves will result in the end of slavery. To be causally efficient, such a recognition would have to be the reason why they chose to hire them. Being aware of the effect of one's action is one thing. Being motivated by the prospect of such an outcome is another.[2] Doctors, to take a case in point, may act in a way they know will lead to the death of a patient without intending to kill the patient. An invisible-hand theorist should only worry about the case where agents not only anticipate the consequence of their actions but also are moved by such anticipation. Someone advancing an invisible-hand explanation should rule out only the possibility that agents act in full knowledge of these consequences without intending to bring about these consequences. As an advocate of parsimony, that person's explanation should minimize only the number of causal

relations that are necessary to explain how the world works. Because 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.

In sum, requiring blindness to the consequence of their actions is a constraint that is unnecessarily restrictive.[3] Agents who are led by the invisible hand shall be protected only against having the outcome that figures among their reasons for acting,[4] and the standard account shall be revised accordingly:

> Revised account (RA)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.

## The case of wayward causal chains

What constraint must an agent meet in order 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, leaving 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.[5] These nations, as Marx recalls, were under the influence of mercantilism, which is the view that prosperity depends on the possession of gold. However faulty mercantilism might be as a theory, it did bring wealth to the nations who applied it, according to Marx. It brought them wealth by motivating them to engage in efforts – to colonize new countries, thus contributing to the proliferation of new needs, 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.

The mechanism at play in Marx's explanation is similar to the one exemplified in La Fontaine's fable *The Labourer and His Children* (cf. Elster, 1987). The fable tells the story of three sons who were too lazy to work in the fields, as their father wished them to. The father told them that there was a treasure buried in the ground. Eager to get rich in a hurry, they turned over the soil in an unsuccessful search for the treasure, and in doing so made it so fertile that they did indeed become wealthy, 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 become so. Getting rich would have been something that they intentionally did if the sons had become rich by way of finding the treasure they were looking for. But wealth came to them through a very different path.

In both La Fontaine's fable and Marx's explanation, the recognition of the effect clearly is central to the reasons why agents act the way they do. In both examples, what occurs corresponds to what was expected to occur. However, the mechanism that is responsible for the correspondence between the two is completely unanticipated. These

examples show that the intention towards P (e.g. to become wealthy) is not yet a sufficient condition for intentionally becoming P (e.g. wealth). The content of one's intention, e.g. to become wealthy, may represent the state of affairs in which that intention is satisfied. Yet having such an intention does not ensure that the state of affairs has been intentionally brought about. This is because the identity between the goal and the produced state of affairs might be a twist of fate. The wealth of 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.

Agents may succeed in bringing about the consequence they intended to bring about, but not in the way they thought the consequence would occur. The consequence is the result of a so-called wayward process 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. Such a wayward process is, aside from the more typical cases, another way in which the invisible hand can be operative. Nozick has this type of case 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' (1994: 314, fn 2). I propose that we allow such cases to be genuine, although less typical, cases of invisible-hand explanations. Although agents led by the invisible hand are in standard cases blind to the end result of the process in which they participate, they may alternatively be blind to the process by which such an outcome is brought about. If a wayward causal chain is a way by means of which the invisible hand can operate, we should allow agents to 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:

> RA2: 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.

## The case of Lewis conventions

Suppose now that the consequence did 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, it seems that 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 many countries. There are many ways of explaining it. One is to describe it as the result of a policy. A legislator decides which side will be the driving side and punishes those who violate that rule. Alternatively, the rule can be described as the result of a collective

agreement. The drivers can discuss the matter and decide which side they will drive on, unaided by any third-party coordinator. Either way, the explanation relies on the ability of some designing minds to devise the rule. For this reason, the explanation will not be an invisible-hand explanation. There is, however, an alternative explanation that does not rely on a centralized system of sanction nor on some previous agreement. The explanation Schelling (1960) and David Lewis (1969) offer of conventions is, as I purport to show, of that kind.

Conventions are, in Lewis's view, solutions to coordination problems. Deciding which side of the road should be the driving side, choosing something (e.g. cowrie shells or coconuts) as a medium of exchange, deciding who should call back when a phone conversation has been interrupted (Lewis, 1969: 9), limiting in some way the use of gas as a weapon (Schelling, 1960: 75) are coordination problems. The latter are the sort of problem you have to deal with when you have an interest in doing whatever the other does whereas more than one option could be a solution. There is no outcome by which agents' 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.[6] What matters is that you and I end up choosing the same one.

It is Lewis's view that a coordination problem is resolved when agents succeed in converging on one of the equally valuable options. The capacity to form the right sort of mutual expectations plays, in this respect, a crucial role. Except in England and its former empire, agents expect each other to drive on the right side of the road. Mutual expectations split in turn into first-order expectations, as when I expect you to drive on the right-hand side and vice versa, and higher-level expectations about what you expect me to expect you to do, about what you expect me to expect you to expect me to do, and so on. Schelling also recognizes such infinitive 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: 87).[7]

This being said, the left side would solve the coordination problem as well. So the crucial question is: What makes the option of driving on the right side of the road the option of concordant mutual expectations?

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, either of the two combinations of options – left/left or right/right – allows us to maximize our expected utility. Moreover, rational choice forbids agents to be influenced by the way the options are labelled (Sugden & Zamarrón, 2006). This is because, in rational-choice theory, what truly distinguishes 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 options 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 labelled, rather than on what is essential to it, namely its payoff.

The preceding reasoning is certainly counter-intuitive. Not only is rational-choice theory unhelpful at explaining how we solve coordination dilemmas, it also forbids us to use a device that seems intuitively useful. Undoubtedly, we rely on the way options are labelled in order to have our expectation converge on one of them. We do so because, as Lewis argues, the distinctive connotations that are attached to the description of each option make some of them more salient, allowing them to become the content of our mutual expectations.

A salient option is an option that commands attention, i.e. one that 'stands out from the rest by its uniqueness in some conspicuous respect' (Lewis, 1969: 35). It is, in Schelling's words (1960), a 'focal point' and can be used as such as a coordination cue. For an option to be salient it must be so for many, and believed to be so among those for whom it is so. 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' (Lewis, 1969: 33–34). In the same vein, he also says that a salient option is 'unique in some way the subjects will notice, expect the other to notice, and so on' (1969: 35).

What makes an option salient? A prominent source of saliency is precedence. An option is eye-catching because we reached it last time. Driving on the right side is a salient option because it is what drivers are used to doing.[8] Other factors may however contribute to the salience of an option. The right side is the side of the road on which it also is legal to drive, a part of everyone's habit. What commands attention may thus be what is legal to do. Conspicuousness can, still alternatively, depend[s] on 'analogy, … accidental arrangement, symmetry aesthetic or geometric configuration, casuistic reasoning, and who the parties are and what they know about each other' (Lewis, 1960: 57). While the cultural context will determine why certain options are in particular more striking than others, the very possibility of relying on salience as a way of meeting each other's expectations is always there.

Four features characterize a Lewis convention (1969: 58). First, it is a regularity 'to which everyone conforms'. Regularities that are only legally prescribed, such as the San Diego law forbidding the possession of more than two dogs, are not Lewis conventions. Second, it is a regularity to which 'everyone expects everyone else to conform'. That second condition excludes the general tendency (not) to follow the traffic monitor's tips. So too is the widespread habit of brushing one's teeth ruled out from the class of Lewis conventions inasmuch as the latter is not a regularity to which, thirdly, 'everyone prefers to conform … on condition that the others do'. Finally, these three conditions must, as a final condition, be the content of common knowledge. The fourth condition excludes the possibility where each driver falsely believes that he is the only one acting strategically within a group made of parametric decision-makers. Everyone must assume that the 'I will if you will' condition is everyone's reasoning, not only one's.

## An invisible-hand explanation

Lewis explicitly intended his theory to replace an explanation of conventions that assign a crucial role to either a legislator or to a collective agreement. He says that legislators (or, in the present case, the 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. The punishments operate as an external incentive that outweighs all reasons I may have for acting otherwise. Driving conventions happen to be the result of a 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 fundamental than fines. 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' (1969: 45). It is especially true of driving conventions where staying alive is at stake.

In line with the tradition of the invisible hand, Lewis also aims to provide an explanation that dispenses with collective agreement. To be sure, Lewis grants that all conventions could in principle originate in, and maintain themselves by, a binding agreement. They need not do so, however. 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' (1969: 34). 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 during a drought because the unconditioned inclination everyone has is not to decrease one's consumption. In a case like this, agents need to bind themselves to prevent themselves from being tempted to free ride, and promising is one efficient way to do this. But unlike prisoner-dilemma-type situations, agents' interests are, in a coordination situation, well served when they simply act on mutual expectations. Promises therefore are superfluous and 'an exchange of declarations of present intention' suffices (1969: 34).

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 some 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 forefathers' 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' (1969: 84).

Finally, the condition of interdependency, which, according to Lewis, defines convention, excludes the possibility that they originate in agreements, i.e. promises to act

unconditionally.[9] 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' (1969: 84).[10]

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. External punishments and agreements might very well play a causal role in the actual emergence and maintenance of conventions. What Lewis distinctively shows is that they are nevertheless inessential to their existence. If either a sanction or an agreement were in fact involved in a given case, they would be contingent facts about that convention.

It follows that Lewis's account of conventions clearly satisfies the 'no designer condition' of all invisible-hand explanations.[11] Yet a Lewis convention also clearly lacks the feature of unintendedness. Driving on the right side is not construed as an unintended consequence of actions directed toward other ends. Agents deliberately converge on that option. Lewis's view, in sum, is that the situation of agents when they follow a convention is a situation they deliberately bring about. It is a solution to a coordination dilemma and not a by-product of some other concern. Nor is it a situation that agents 'stumble upon', to use Ferguson's oft-quoted expression.[12] It results instead 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 feature of an invisible-hand consequence. What is more, no wayward causal chain is involved in the establishment of a Lewis convention. A Lewis convention is a consequence of many actions performed by agents who, first, have its establishment in mind; who, second, let its establishment enter into their reasons for acting; and, third, whose intentions are the non-deviant causes of that consequence. Our last revised account of the theory of the invisible hand, RA2, turns out to be as unable as the standard definition to accommodate a Lewis convention. If the latter really illustrates an invisible-hand explanation, any account of the latter should not rule out the case where agents successfully act with the view to producing the outcome to be explained.

Specifying such an account is uneasy, however, for it will have to accommodate the case of Lewis conventions without blurring the distinction between invisible-hand explanations and social-pact explanations. To explain an outcome as the result of a social pact is indeed to depict it as what agents intended to bring about and managed to produce in the way they intended to produce it. Remember, however, that a social-pact account and an invisible-hand explanation are incompatible. The invisible hand denies the need of collective agreement and the latter is a crucial stage of a social pact. So, any account should be able to draw the line between these two conflicting views.

The solution, I believe, is to distinguish between two modes of intention, namely an individualistic mode, an I-intention, and a non-reducible collective mode, a we-intention. While the latter is typically at play when agents collectively agree to conform to a convention, the former is operative when agents can rely neither on speech nor on any explicit bargaining as a way to converge on the same convention.

Searle's example (1990) of the graduate businessmen case presented above can be helpful once again in order to grasp the difference between these two modes of intention. 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' (1990: 407), 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 in 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.[13] They (like any social pattern) can be the result of many actions performed by agents who are acting individually, albeit interdependently, and with the same purpose in mind. Or they 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.

If a Lewis convention can be shown to be one that specifically involves an I-intention, the boundary condition of an invisible-hand explanation will have been found. I indeed propose to draw the line that separates invisible-hand explanations from intentional-design explanations by using the difference between these two modes of intention. On the related account, a convention would be the product of an invisible-hand process only in cases where it arises out of a process during which agents reason and act individually – during which agents are each moved by an I-intention. Let agents reason and act jointly – let the latter be moved by a we-intention – however, and the same convention would fall within the scope of an intentional-design explanation.

> RA3: 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.

## Acting jointly

RA3 is meant to be a compelling way of drawing the line between invisible-hand explanations and intentional-design explanations. On the proposed account, the former does not differ from the latter in virtue of its distinctive ability to describe its explanandum as an unintended consequence, as the standard account suggests. It differs from the latter in virtue of its ability to describe its explanandum as the intended result of many individual (rather than joint) actions.

RA3 will sound compelling if the example that motivated it is itself convincing. In other words, it is only if a Lewis convention is persuasively presented as involving agents who do not act jointly that RA3 will be accepted. Yet the latter claim may be resisted. Sceptics may indeed quote 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.

They will select the sentence where salience reasoning is said to require a 'meeting of minds' (Schelling, 1960: 163) without which 'some kind of collaborative or mutual accommodation' (1960: 83) can hardly be achieved. They will stress the passages where Lewis says that the salient option will be found 'by putting ourselves in the other fellow's shoes' (1969: 27). Surely, these quotes do not serve our purpose well, to say the least. They admittedly suggest that salience reasoning is a matter of attuning oneself to the others' minds – of 'doing one's best to mirror each other, mirror each other mirroring each other, and so on', as Lewis also says (1969: 32), in a way that seems typical of joint actions. A focal point, in other words, seems to be both discovered by a we-attention and issuing in a we-intention to act on its basis.

This section is devoted to making the contrary interpretation plausible. Notwithstanding the various quotes in the previous paragraph, there are reasons to believe that a Lewis convention involves no we-ness. A straightforward way of advancing the point is to recall its underlying game-theoretical framework. Because game theory postulates that no one other than individuals can choose, the question 'what should we choose?' does not make sense to any of its advocates. The reason is that individuals, the latter assume, are the only unit of agency conceivable.[14] Speaking about a group mind can only be, at best, a *façon de parler*.

Still, the idea of an irreducible plural subject is not the only way to approach joint intentions. Another way – one that is compatible with the game theorist's stance against group minds – is to assign it to the content of the intentions, rather than to their allegedly unique bearer.[15] On the related view, taking the group's perspective affects what agents intend to do. Such a content-based approach to joint action comes in many versions, some of which could apparently accommodate the sort of agency that gives rise to a Lewis convention. Next I critically review these versions and advance some reason to be sceptical about their relevance. My purpose is quite humble, suitably couched in terms of casting doubts on – rather than refuting – the operation of a collective intention in a Lewis convention. My strategy is as follows: to review the various features of a Lewis convention in light of which it may appear to be based on a we-intention and show how each of these features is subject to counter-examples of two kinds. The first are examples of individual actions that display those features, and the second are examples of we-actions that lack them.

To begin with, take an account of joint action that conceives it as nothing more than several infinitely reiterated I-intentions, plus common knowledge of the situation. On such an account, we intend to do something together if and only if:

  – I intend to F and you intend to do something.
  – I know that you intend to F and you know that I intend to something.
  – You know that I know that you intend to do something and I know that you know that I intend to do something, and so on.

Surely a Lewis convention is compatible with such a distributive account of joint action. However, the latter can be rejected on the basis of its inclusion of other cases, such as Searle's example of the businessmen graduate students, which are clearly not cases of joint actions.

The relevance of approaching a Lewis convention as involving a capacity to act together may, however, be defended by revising the above account. The amendment consists in adding the disposition to act interdependently (see Bratman, 1993; Miller, 2001). While the businessmen clearly lack such a disposition, since they have learned to pursue their individual interest regardless of what the others do, agents involved in a Lewis convention clearly have it. Interdependency is also at the core of the various strategic decisions on which a Lewis convention is based. I intend to choose the right side conditionally upon my knowing that you intend to choose the same side, and vice versa. 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...?' (Schelling, 1960: 54). To make room for a Lewis convention, one thus only needs to amend the previous account of joint action with the additional and mutually known 'I will if you will' condition.

But is interdependency really an essential ingredient of all joint actions? There are various reasons to consider it is too weak a requirement. As Michael Bacharach has shown (see Bacharach, 2006; also Gold & Sugden, 2007), interdependency is, with common knowledge, a feature of all Nash equilibria, including some that are intuitively not the outcome of a collective intention. The case of mutual defection in the prisoner's dilemma is such a case. Even if each individual's intention to defect is a best response to her true beliefs about the others' action, the resulting defection certainly does not result from a we-intention. Another counter-example is the case of Susan and Bill, the spiteful theatregoers, which Kay Mathiesen puts forward (Mathiesen, 2002) in order to show why interdependency is not a sufficient feature of joint action. 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 Susan 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 a Salesman*. Yet 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, and yet they may be unconnected by any we-intentions to act together. Interdependency is therefore not a sufficient condition of joint actions.

Nor is it, as Mathiesen 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 Mathiesen 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. Interdependency, i.e. the rule 'I will if you will', which partly defines a Lewis convention, thus seems not to be a feature of the way conventions are jointly discovered and followed.

At this point, those who remain convinced that Lewis couched his theory of conventions in a way that assigns a role to collective agency could invoke the participatory view that characterises many accounts of the latter. They could say that, when drivers choose

the right side of the road, they have a sense of what they do as 'doing their part' of a combination of action. Such a personal contribution to an action that has other contributors is considered an essential component of joint actions. The latter are typically performed by several agents who see themselves as doing their bit and expect each other to do so. The requirement is at the core of Sugden's idea of collective intentionality, which he articulates in terms of team reasoning. When agents intend to act together, they view themselves as the members of a team, Sugden argues, and consider their individual contribution as a way of 'playing their part' in the combination of actions that is best for the team (Sugden, 2000: 187; 2003: 167–168; Sugden & Gold, 2007: 119; Sugden & Zamarrón, 2006: 615). This participatory view of joint action is also central to Searle's conception. The individual participations, as he also stresses (Searle, 1990), derive from the joint action. To be able to see what I do as 'doing my bits of act A', I must first see what I do as 'we are doing act A'.

Now it is true that agents acting individually may also act by bits.[16] They may partition an action into smaller, more manageable parts. And they may also see the latter as parts of an overall act that they intend to perform. Yet these sub-acts are, in the case of an individual action, performed by the same agent. Let someone else take charge of a few of these sub-acts,[17] in other words, and the latter will become a case of joint action.

A Lewis convention also exhibits the same participatory structure. Driving on the right side can be described as a choice drivers make collectively, by each of them doing their part in the combination right/right. A Lewis convention thus involves agents who view their choice in a participative way. The coordination dilemma will not get solved, each agent knows, if each fails to do 'his part of one of the possible coordination equilibria' (Lewis, 1969: 25).

But is the sense of doing one's part an ingredient that is essentially constitutive of a joint action or is it a mere desideratum of the latter? Such a requirement may be too restrictive, as Hans-Bernard Schmid argued, in light of the possibility of having 'negligent, sloppy, unfocused, forgetful and weak-willed' (Schmid, 2009:47) and otherwise 'recalcitrant' (2009: 47) participants in actions that nonetheless remain jointly performed. The participatory requirement that characterizes the dominant view may be founded on a 'bias for smooth cooperation' that makes us blind to the 'unruly fellows who do not only just happen to fail but even have the intention not to do their part in a shared cooperative activity?' Schmid puts forward the case of two persons who have the intention to lunch together at the cafeteria but are so engaged in the topic of their lively debate that one of them ends up forgetting about where she is going. In spite of the lack of participatory intentionality, it remains appropriate to refer to their action as going to the cafeteria. Such a case certainly casts doubt on the necessity of actively doing one's part in an action as a condition of joint action. It consequently questions whether agents who succeed in converging on a Lewis convention really are acting together.

The idea that agents who are involved in forming a Lewis convention act individually – even if they act interdependently and are willing to do their part and have 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 jointly acting with the shared purpose of bringing them about together. One well-known instance of such an

account is Margaret Gilbert's idea of convention (Gilbert, 1989, 2008), which self-consciously differs from Lewis's in many respects. First, it is an account that rules out all of Lewis's conditions. For example, the convention that the members of some departments adopt regarding formal clothing at department meetings, she argues, shows why conformity, expectation and an underlying coordination structure are neither sufficient nor necessary. Second, unlike Lewis's individualistic account, Gilbert's is said to be a holistic one in that it appeals to a non-reducible concept of joint commitment. Whereas, in the case of personal commitments, 'each must commit himself', in the case of joint commitment 'all commit all' (Gilbert, 2008: 11). A joint commitment arises when each party tacitly or explicitly expresses to the others his willingness 'to participate with the others in committing them all' (2008: 11). Those who create a joint commitment together impose a constraint on each of the parties with respect to what it is rationally open to them to do. Each party, as Gilbert says, owes to every other party his conformity to the commitment. A third un-Lewisian feature of Gilbert's account is the offence which someone's failure to conform will induce in those who have fulfilled their commitment. Such an offence opens the way to a fourth feature, which is the position in which the offended find themselves able to rebuke the transgressor for not having performed his committed action. The three ingredients of joint commitment, offence and reprimand confer on conventions a think-normative dimension which Lewis's account certainly lacks, insofar as he thinly conceives the 'oughtness' of conventions as a matter of instrumental rationality (Lewis, 1969: 97–100).

In sum, because it requires more than mere interdependency and common knowledge, and because it involves a robust social normativity, Gilbert's conception of conventions as the result of joint action adequately serves as a limiting case for the operation of the invisible hand. It points to what agents should not do and think for being led by it.

The performance of joint action represents 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.[18] In the search for such a line, several explanatory options have been encountered, which it is time to review. These conflicting or partly overlapping mechanisms by which social outcomes arise out of agents' coordinated choices differ from one another in virtue of the sort of coordinating clues that they put out. Choices, we have seen, can be coordinated as the result of: (i) the intervention of a central authority, (ii) a social pact (Locke), (iii) a wayward causal chain (Marx), (iv) mutual expectations around some focal points (Lewis), (v) team reasoning (Sugden) or as the product of (vi) some previous joint commitment (Gilbert).

## Conclusion

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 ensure 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.

The proposed revised account 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'. It is superior to the standard account in three ways. First, it saves us the task of resolving the various puzzles that surround the identification of unintended consequences. Second, it accommodates a case of invisible-hand explanations, namely all Lewis conventions, which would have otherwise been ruled out. Finally, it distinguishes the invisible hand from the social pact by assigning a different kind of agency to each of these two rival accounts of social reality.

Nevertheless, it may still be found important to keep the unintendedness assumption. It may, for example, be justified on the ground that it reflects the condition of opacity in which agents find themselves in regard to the social reality. It may alternatively be of instrumental value inasmuch as it enables the advancement of a fundamental, that is non-circular, explanation, one that will not refer to the outcome to be explained in the course of the explanans (Nozick, 1974). Or it may simply be praised for the quite elitist pleasure of giving a description of a social phenomenon that the very agents who make it possible cannot themselves provide.

## Acknowledgements

## Funding

## Notes

1.  Millar was writing in 1771, at a time when the issue of slave emancipation was widely discussed but had not been resolved in most European countries (let alone in America). This might explain the historical inaccuracy of his explanation, since he could not foresee that slavery would be terminated by state intervention rather than by spontaneous economic processes. He therefore was simply making a conjecture as to how things might develop, not on what had happened. Yet it is not certain, one may alternatively argue, that, had Millar correctly anticipated or observed the abolition of slavery, he would have revised his explanation accordingly. His invisible-hand explanation may indeed be deliberately a 'potential' one (Nozick, 1974: 7), and an illustration of Dugald Stewart's (1753–1828) idea of 'conjectural histories' (Stewart, 1858: 34). (D Stewart's [1858] *Biographical memoirs of: Adam Smith,*

LLD, William Robertson, DD and of Thomas Reid, DD, read before the Royal Society of Edinburgh, were collected into one volume, with some additional notes, in *Collected Works*, vol. X, Edinburgh, 1858.)

2. The same revision has been proposed 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 phenomenon, must be unrecognized 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 recognised 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: 119).

3. 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.

4. 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 recognize 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 insofar as it ensures the behaviour is not intended in order to produce the effect to be explained.

5. The explanation appears in Marx (1973) and is presented by Elster (1985: 22–24).

6. There is a caveat to this claim. Although these two options are commonly considered as equally attractive, driving on the right side actually is an inferior option because people have the tendency to swerve to the left in an accident, with the prospect of running into oncoming drivers. If this is so, the dilemma involved in choosing between left and right should therefore be construed as a Hi-Lo game rather than as pure coordination dilemma.

7. Lewis acknowledges a practical limit to the length of that spiral, however, namely the fourth level.

8. Using an evolutionary game-dynamic approach, Young (1996) offers an account of the emergence of the right-hand driving rule. He shows that, although the left side was the dominant convention in most of Western Europe before the French Revolution, a chain of historical circumstances, symbolic reasons and rational inclination to adjust to the nearby countries' rule tipped the balance in favour of the right side.

9. Surely, we may promise to conform to R on condition that others do as well, as Lewis observes (1969: 84). However, most conceptions that derive the existence of conventions from an agreement nevertheless do not conceive the promises to conform to R to have such a conditional character.

10. 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's idea of convention.

11. This classification is not uncommon. Having distinguished between 'straightforward and more complicated cases' of invisible-hand explanations, Brennan and Pettit subsume Lewis's account of conventions under the latter category (Brennan & Pettit, 1993: 203–207; see also Aydinonat, 2008: 147–152).

12. Here is the full quote: 'Nations stumble upon establishments, which are indeed the result of human action, but not the execution of any human design' (Ferguson, 2006[1782]: 187).

13. Besides these two approaches, the French school of 'Economies des conventions' represents still another conception of conventions (cf. Dupuy et al., 1989; Orléan, 2004). Unlike the Lewisian account, the members of the French school will stress the condition of uncertainty in which agents find themselves as the result of their bounded rationally. They will also cast light on the agents' interpretative skills as well as on the normative dimension which gives to conventions either their legitimate or, on the contrary, their unjustified character. Unlike the holistic approach, however, the French school spells out its conception of conventions in individualistic terms. The French tradition therefore represents an interesting alternative view, one that stands somewhere in between the two conflicting views to which I refer in the article. Because it is an individualistic approach, however, I find it too close to the Lewisian account to play the role of the limiting case, vis-a-vis the invisible hand, which the Gilbertian approach suitably illustrates.

14. The idea that no other entity but individual agents can choose is an ontological presupposition – an assumption – about the basic constituents of social reality (cf. Mäki, 2001: 369–372), which is, as such, mostly left un-supported by game theorists and mainly debated by philosophers working in the field of social ontology.

15. A third approach conceives the difference between joint and individual actions as involving a distinctive mode of representation. On this view, there are two modes of intending one and the same thing, 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 F', rejecting the content-based approach of the form 'I intend that we F', on the one hand, and an implausible world spirit floating above individuals, on the other.

16. '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 a 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: 167).

17. There are, interestingly, an indeterminate number of bit(s) of my individual action that I can delegate to someone else before turning it into a joint action.

18. The difference between the two sorts of accounts is where the dividing line between invisible-hand explanations and intentional-design explanations lies. Compatible with that claim, it is important to stress, is any appraisal regarding the merit of each sort of explanation.

## References

Aydinonat E (2008) *The Invisible Hand in Economics: How economists explain unintended social consequences*. London/New York: Routledge.

Bacharach M (2006) *Beyond Individual Choice: Teams and frames in games theory*, edited by Gold N & Sugden R (eds). Princeton, NJ: Princeton University Press.

Bratman M E (1993) Shared intention. *Ethics* 104(1): 97–113.

Brennan G, Pettit P (1993) Hands invisible and intangible. *Synthèse* 94(2): 191–225.

Cowen T (1998) Do economics use social mechanisms to explain social mechanisms? An analytical approach to social theory. In: Hedström P, Swedberg R (eds) *Social Mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press, 125–146.

Dupuy J-P, et al. (1989) Introduction. *Revue Economique* 40(2): 141–145.

Elster J (1978) *Logic and Society*. Chichester: Wiley.

Elster J (1985) *Making Sense of Marx*. Cambridge: Cambridge University Press.

Elster J (1987) *Le Laboureur et Ses Enfants. Deux essais sur les limites de la rationalité*. Paris: Editions de Minuit.

Ferguson A (2006[1782, 1767]) *An Essay on the History of Civil Society*. London: T Cadell.

Gaus G (2011) Explanation, justification, and emergent properties: An essay on Nozickean metatheory. In: Bader RM, Meadowcroft J (eds) *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*. Cambridge: Cambridge University Press, 116–144.

Gilbert M (1989) *On Social Facts*. London/New York: Routledge.

Gilbert M (2008) Social convention revisited. *Topoi* (27): 5–16.

Gold N, Sugden R (2007) Collective intentions and team agency. *Journal of Philosophy* 104(3):109–137.

Grimen H (1994) Causally inefficient knowledge and functional explanation. *Social Science Information* 33(1): 117–127.

Hayek FA von (1948) *Individualism and Economic Order*. Chicago, IL: University of Chicago Press.

Lewis D (1969) *Convention*. Cambridge, MA: Harvard University Press.

Mäki U (1990) Practical syllogism, entrepreneurship and the invisible hand. In: Lavoie D (ed.) *A Critique of the Analytic Hermeneutics of G. H. von Wright: Economics and hermeneutics*. London: Routledge, 149–176.

Mäki U (2001) The way the world works (www): Towards an ontology of theory choice. In: Mäki U (ed.) *The Economic World View: Studies in the ontology of economics*. Cambridge, Cambridge University Press, 369–389.

Marx K (1973[1857–1861]) *Grundrisse der Kritik der Politischen Ökonomie* (Outlines of the Critique of Political Economy), transl. by Nicolaus M. London: Penguin.

Mathiesen K (2002) Searle, collective intentions, and individualism. In: Meggle G (ed.) *Social Facts and Collective Intentionality*. Frankfurt: Dr Hänsel-Hohenhausen AG, 185–204.

Menger C (1892) On the origins of money. *Economic Journal* 2(6): 239–255.

Millar J (2006)[1771]) *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*, ed. by Garrett A. Indianapolis, IN: Liberty Fund.

Miller S (2001) *Social Action: A teleological account*. New York, NY: Cambridge University Press.

Nozick R (1974) *Anarchy, State and Utopia*. Oxford: Blackwell, Basic Books.

Nozick R (1994) Invisible-hand explanations. *American Economic Review* 84(2): 327–332.

Orléan A (2004) L'Economie des conventions: Définitions et résultats. In: Orléan A (ed.) *Analyse Economique des Conventions*. Paris: Presses Universitaires de France. (2nd edn)

Petsoulas C (2001) *Hayek's Liberalism and Its Origins: His idea of spontaneous order and the Scottish enlightenment*. Studies in Social and Political Thought. London: Routledge.

Schelling T C (1960) *The Strategy of Conflict*. Cambridge, MA: London: Harvard University Press.

Schelling T C (1969) Models of segregation. *American Economic Review* 59(2): 488–493.

Schmid H-B (2009) *Plural Action: Between philosophy and the social sciences*. Dordrecht: Springer Verlag.

Searle JR (1990) Collective intentions and actions. In: Cohen P, Morgan J, Pollack M (eds) *Intentions in Communication*. Cambridge, MA: MIT Press, 401–415.

Steiner H (1978) Can a social contract be signed by an invisible hand? In: Birnbaum P, et al. (eds) *Democracy, Consensus and Social Contract*. London/Beverly Hills: SAGE Publications.

Sugden R (2000) Team preferences. *Economics and Philosophy* 16(2): 175–204.

Sugden R (2003) The logic of team reasoning. *Philosophical Explorations* 6(3): 165–181.

Sugden R, Gold N (2007) Collective intentions and team agency. *Journal of Philosophy* 104(3): 109–137.

Sugden R, Zamarrón I (2006) Finding the key: The riddle of focal points. *Journal of Economic Psychology* 27: 609–621.

Tuomela R (1984) *A Theory of Social Action*. Dordrecht: Reidel.

Ullmann-Margalit E (1978) Invisible-hand explanations. *Synthese* (39): 263–291.

Young P (1996) The economics of conventions. *Journal of Economic Perspectives* 10(2): 105–122.

Zaibert L (2004) Toward meta-politics. *Quarterly Journal of Austrian Economics* 7(4): 113–128.

## Author biography

**Emma Tieffenbach** graduated in history from the University of Geneva (1998) and obtained a Diplôme d'études approfondies (2000) at the Centre de Recherche Politique Raymond Aron (EHESS, Paris). Her doctoral dissertation (University of Geneva) deals with the cognitive value of 'invisible-hand explanations'. Since April 2012, she has been working on the nature and motivational power of economic values at the Swiss Center for Affective Sciences (University of Geneva). She was awarded a Marie Heim-Vögtlin subsidy from the Swiss National Foundation.