EDNA ULLMANN-MARGALIT

INVISIBLE-HAND EXPLANATIONS*

1. THE SETUP

A long-standing tradition makes the first step towards introducing order into the plethora of the phenomena encountered by man by classifying them as either 'natural' or 'artificial'. Obvious and sensible as this classification may appear to be, it is deficient. Characterizing 'natural' phenomena as those that are wholly independent of the human sphere and 'artificial' phenomena as those that involve human agency leaves the deficiency at least partially concealed. It is brought into relief, however, once it is said that the realm of the 'natural' consists of everything that is the result neither of human action nor of human design, while the realm of the 'artificial' of everything that is the result of both: what is clearly seen to be lacking is the at least hypothetical realm of everything that is the result of human action but not of human design.

The notion that this middle realm is real rather than hypothetical is also hardly new[1], as is well attested by its articulation in the writings of the 18th-century Scottish school. In order to indicate its weightiness, and for suggestive purposes too, I shall refer to it as the realm of the *social*. I hasten to add, though, that this terminology should not be taken to imply that all social patterns and institutions belong here: some are, of course, the product of human execution of a human plan, and as such belong to the 'artificial' category.[2]

Those social patterns that can be viewed as results of human action but not of human design are candidates for a special kind of explanation which, following Nozick who takes up Adam Smith's cue, will be called *invisible-hand explanations.*

2. THREE PARADIGMATIC EXAMPLES

An invisible-hand explanation, very roughly, is one that explains "what looks to be the product of someone's intentional design, as not being brought about

Synthese 39 (1978) 263–291. *All Rights Reserved.*
Copyright © 1978 by D. Reidel Publishing Company, Dordrecht, Holland.

by anyone's intentions". (Nozick [16], p. 19.) Before attempting to fill in this rather meager formulation, let us pin down our imagination and ideas with the help of a few examples.

(i) The continuous creation of money within the banking system: No one needed to have invented the commercial banking system, nor need anyone have intended it to function so as to continuously create money. The usual story that accounts for both begins with the early goldsmiths who used to be paid a small fee for the safekeeping of people's gold and valuables. It proceeds with those intelligent goldsmiths who came to realize, first, that they don't necessarily have to give back to the customer exactly the same piece of gold that he had deposited, and, later, that since not all deposits are withdrawn together and new deposits tend to balance withdrawals, only a small percentage of the cash entrusted to them is needed in the form of vault cash. The rest of the story has to do with these shrewd bankers' investment in securities and loans of most of the money deposited with them, leading to the account of the actual creation of money through the consideration of the overall impact of this newly-developed banking system as a whole rather than of each small establishment taken in isolation.[3]

(ii) Nozick's account of the rise of the so-called ultraminimal state: This account is consciously styled as an invisible-hand explanation (Nozick [16], Part I). It begins with a Locke-like state of nature and lists the problems concerning the enforcement of individuals' rights therein. It then proceeds in successive stages through the formation of the naïve and inefficient mutual-protection associations, each member of which being liable at any time to be called upon to assist any of the other members in defending or enforcing his rights, then of the more professional private protective agencies, where there is division of labor and exchange and some people are hired to perform protective functions, and finally of the single protective agency that emerges as dominating a geographical area. The rest of the story has to do with supplying arguments as to why this dominant protective agency, which appears to fall short of being a state, actually does constitute an ('ultra-minimal') state.

The invisible-hand process involved is compressed by Nozick into the following formulation:

> Out of anarchy, pressed by spontaneous groupings, mutual-protection associations, division of labor, market pressures, economies of scale, and rational self-interest there arises something very much resembling a minimal state - -. (pp. 16–17)

(iii) The development of media of exchange: In the light of Ludwig von Mises' account (von Mises [15], pp. 30–34) neither a decree issued by an enlightened ruler nor explicit agreements fixing a medium of exchange need be posited. The obvious inconveniences attending the direct barter system, and the obvious value of something like gold – which, in addition to being desirable to begin with, is also scarce, divisible, non-perishable, portable, immutable, etc. – are sufficient in order to spell out the details of the invisible-hand account in question.[4]

In all of these cases the phenomenon explained is shown to be the product neither of centralized decisions nor of explicit agreements to bring it about; rather, it is presented as the end result of a certain process that aggregates the separate and 'innocent' actions of numerous and dispersed individuals into an overall pattern which is the very phenomenon we set out to account for.

But what is it that typifies that which qualifies for an invisible-hand explanation? And why does it initially look as if it is the implementation of someone's intentional design? And if it is indeed shown not to have been brought about by anyone's execution of a plan, just how *is* it claimed to be brought about?

3. ELIGIBILITY FOR INVISIBLE-HAND EXPLANATIONS

Suppose we encounter a crater so regular that we naturally assume it to have been carefully dug by some people according to a detailed plan. Suppose further that this crater is used as a water reservoir for which purpose, we realize, it is ideally located as well as ideally shaped. This fact of course increases our confidence in our original assumption. But now suppose some geologists chance along and convince us that this crater was actually formed as a result of volcanic action, its regular shape being satisfactorily accounted for by the application of an elaborate piece of stratigraphy to that area. Have the geologists provided us with an invisible-hand explanation of the formation of the crater?

Bearing in mind the introductory remarks of Section 1, I propose to answer this question in the negative. The geologists have replaced an explanation of the crater-as-an-artifact with an explanation of the crater-as-a-natural-phenomenon. What I want to restrict the expression 'invisible-hand explanation' to, on the other hand, are social phenomena. Moreover, in the course of explaining these social phenomena it is essential, I suggest, that not

only the explanation in terms of human design, which 'naturally suggests itself', but also the (invisible-hand) explanation that purports to replace it turn on human agency. Once again: it is with the unintended consequences of human action that we are here concerned.

At the same time, however, we would certainly want to legislate that not every unintended consequence of human action qualify as an invisible-hand explanandum. Just think of the numerous occasions on which our children come to us in tears begging to be excused for the latest calamity they have wrought on the ground that they "didn't mean to". Quite obviously, the accidental by-products of many of our actions, even of several individuals' actions taken together, hardly constitute a promising recruiting ground for invisible-hand explanations. This holds true even if an initially plausible alternative explanation in terms of intentional design is indeed available in all these cases.

Or consider the case of someone who has a flat tyre while driving through the busy morning street. A "hidden-hand" explanation (see Nozick [16], p. 19) of this mishap would be that someone spread broken glass there with just this purpose in mind. With some additional paraphernalia this could easily turn into a full-blown "conspiracy" explanation (see Popper [17], pp. 94–95) – especially if we were to begin the story with seven punctured tyres that same morning rather than one, and if we mentioned the existence of a service station just a hundred yards down that street. But suppose it were established later that the broken glass was there completely 'by accident' – that, say, a delivery boy with a crateful of bottles had innocently tripped there earlier: would this amount to an invisible-hand explanation of the seven flat tyres (or of seven late arrivals to work that morning, for that matter)?

The answer to this question, I submit, cannot be expected to be grounded in theoretical considerations. We face, once again, a practical issue involving a decision: given that the flat-tyres case is one in which – as required – an explanation in terms of 'human accident' replaces an explanation in terms of human intentional design, do we actually *want* it to count as having been explained invisible-handedly? My inclination is to answer this question with a No. While not attempting to defend this answer as *true*, I shall proceed to argue for it from the consideration of the explanatory import we expect invisible-hand explanations to have.

The way I see it, it is not the fact that an explanation in terms of

unintended consequences of human action can successfully replace an initially plausible 'hidden-hand', or 'conspiracy', explanation that should be taken to bear the explanatory brunt of invisible-hand explanations. Attention should rather be directed to the fact that it is only when the social pattern or institution to be explained has a *structure* beyond a certain degree of complexity that the invisible-hand explanation of it has a point. Furthermore, I suggest that the explanatory interest of an invisible-hand explanation increases with the extent to which there is a difference in type between the overall pattern to be explained and the individual actions which are supposed to bring it about, as well as with the complexity of the intermediary process. Thus, when the aggregate result of individual actions, albeit unintended, is merely an *amplification* of these separate actions I don't quite see the point in saying that the explanation of the resultant phenomenon as the unintended (accidental) consequence of the individual actions – even when many individuals are involved –constitutes an invisible-hand explanation thereof. Think of the loud noise, in a large and crowded hall, produced unwittingly by everyone's conversing in their normal tone of voice.

To be sure, 'structure', 'difference in type', and 'degree of complexity' are hardly the most precise of expressions; nor can they be expected to be made precise in this context. But it seems to me that when conjoined with the paradigmatic as well as with the rejected examples presented so far these terms do indicate the direction in which the explanatory import of invisible-hand explanations lies. They also suggest, what I take to be basically true, that the explanatory import of invisible-hand explanations is a matter of degree.

To recapitulate: An invisible-hand explanation explains a well-structured social pattern or institution. It typically replaces an easily forthcoming and initially plausible explanation according to which the explanandum phenomenon is the product of intentional design with a rival account according to which it is brought about through a process involving the separate actions of many individuals who are supposed to be minding their own business unaware of and a fortiori not intending to produce the ultimate overall outcome.

It seems to me to be quite clear at this point that the onus of the explanation lies on the process, or mechanism, that aggregates the dispersed individual actions into the patterned outcome: it is the degree to which this mechanism is explicit, complex, sophisticated – and, indeed, in a sense

unexpected – that determines the success and interest of the invisible-hand explanation in question.

4. WHENCE THE APPEARANCE OF INTENTIONAL DESIGN?

Before saying more on the invisible-hand process I would like to turn to the question of why it is that the invisible-handedly explained phenomenon 'looks like' the product of someone's intentional design. That is, why is it that an explanation in terms of the execution of a premeditated plan naturally suggests itself in the cases we are considering, and why is it that it has considerable initial plausibility? An answer to this, I surmise, will also bear on the further question of whether the availability and naturalness of such an intentional-design explanation are essential in order for the rival invisible-hand explanation to count as such, or is it merely the case that they usually happen to be concomitant features of such explanations.

By way of addressing these issues, let us put on record the observation that, when called upon to explain human actions and their results, we usually tend to do this in terms of the beliefs, intentions, and goals of a designing agent. All the more so when the phenomenon to be explained is patterned: when it exhibits structure, order, design. Now this tendency is usually successful and hence amply reinforced, since most patterned results of human action really are artifacts produced through intentional design.

Indeed, the manifestation of orderliness and apparent design in the physical universe was taken by a venerable tradition of thinkers to be the result of superhuman action, thereby suggesting – nay, proving – the existence of a superhuman, consummate designer. I am referring of course to the physicotheological tradition that stretches from Plato and Aquinas to Boyle, Ray and Newton, Locke, Clarke, Butler and Paley, and down to J. C. Maxwell. It took the powerful minds – and all the logical arsenal at their disposal – of Hume and Kant, as well as the works of Darwin and Mill, to explode the logic of this Argument from Design, even while acknowledging its deep-rooted prepossessions and its extraordinary psychological grip.

There exists, then, this bias, which we may perhaps term the 'artificer bias', that leads us to postulate a designer whenever we encounter what looks like evidence of orderliness and patterned structure – whether the model be that of the artist creating works of beauty or that of the inventor and

manufacturer of elaborate machines. In addition to these, there is yet another model which may be assumed to corroborate the tendency produced by the artificer bias in the cases we are here concerned with. What I have in mind is the salient and invariable connection, within the realm of living organisms, between the manifestation of coordinated activity on the one hand and the existence of a central nervous system on the other. (See Cummins [4], p. 748.) Now, the phenomena to be invisible-handedly explained, where many individuals are involved, each doing his bit in such a way that an overall structured pattern is produced, are certainly likely to look like cases of coordinated activity. Hence the rather natural 'inference' to the existence of a central planning body that governs this dispersed activity and directs it in accordance with its intended scheme.[5]

A further consideration in this connection is that often the explanandum phenomenon performs a valuable *function*. (Thus, currency as a medium of exchange usefully alleviates the intolerable burden of finding the 'double-coincidence of wants' required for direct barter exchanges.) I shall have more to say about functions below. In the meantime I just wish to point out that when the overall pattern produced is not only well-structured but is also functional, then both the picture of the artificer manufacturing machines and the picture of a central command center controlling and coordinating dispersed activity apply. As a result the tendency to 'naturally' postulate an intentional designer in such cases is of course enhanced.[6]

The considerations dwelt upon so far show, I take it, why it is that a phenomenon eligible for an invisible-hand explanation is naturally and readily explained in terms of intentional design. Moreover, from the nature of these considerations it can be seen that the more structured the pattern to be explained is, the more it exhibits orderliness and design – and in certain cases function as well – the more likely it is that an explanation in terms of intentional design will naturally suggest itself and will look considerably plausible. At the same time, however, as has been indicated earlier, the more structured and complex the pattern, the greater the challenge it poses to whoever proposes to explain it invisible-handedly, and the more potentially successful and satisfying the explanation itself. So the naturalness and availability of the intentional-design explanation are not mere contingent concomitant features of an invisible-hand explanation worthy of the name: they are part and parcel of its very nature.

In the preceding discussion I have several times used the expression 'readily available and initially plausible' to characterize the explanation in terms of intentional design which the invisible-hand explanation purports to replace. This use was overly innocent and deserves comment. By way of unpacking it I would like to suggest that this intentional-design explanation acquires its initial plausibility precisely because it is readily available, and furthermore that it is readily available because of the saliency and grip of the two 'pictures' (– the machine manufacturer model and the command-center model –) discussed above. As to the *probability* of the intentional-design explanation vis-à-vis its invisible-hand rival, this is an issue I should like to come back to below, when discussing the question of the *cogency* of invisible-hand explanations.

5. THE INVISIBLE-HAND PROCESS

Let us now pick up the thread left dangling at the end of Section 3 and focus our attention on the invisible-hand process. By 'the invisible-hand process' is meant the aggregate mechanism which takes as 'input' the dispersed actions of the participating individuals and produces as 'output' the overall social pattern. As was claimed earlier, it is this process which bears the explanatory brunt of invisible-hand explanations. When saying more on this it will be useful to keep in mind the three paradigmatic examples offered in Section 2, to which we may now add the classical account, due to Adam Smith, of the equilibrial pricing system that develops within the perfectly competitive market.[7]

As clearly emerges from these examples, the invisible-hand process is typically conveyed by means of a 'story'. Furthermore, this story usually consists of successive stages, the last of which being the pattern whose explanation is sought. As such, the full-fledged description of the invisible-hand process falls under Hempel's category of *genetic explanations* which, as he says, "present the phenomenon under study as a final stage of a developmental sequence, and accordingly account for the phenomenon by describing the successive stages of that sequence" (Hempel [11], p. 447).[8]

It is my view that to look for generalizations over these stories, or to seek to unearth 'the logic' of the processes, would be a futile misplacement of the desideratum. The path that does seem to me to be worth pursuing leads in the

direction of the *constraints* we assume to be – or want to see – imposed on such stories if they are to form the backbone of good invisible-hand explanations.

The first constraint I shall propose is that which makes invisible-hand explanations attractive to methodological individualists. It stipulates that the description of the initial stage from which the process is supposed to take off is to consist of nothing but the private intentions, beliefs, goals, and actions of the participating individuals, in a specified setup of circumstances. Moreover, it is laid down that these individuals do not have the overall pattern that is ultimately produced in mind, neither on the level of intentions nor even on the level of foresight or awareness.[9] The basic picture underlying invisible-hand explanations, then, is that of a bird's eye view that encompasses numerous individuals, each busily doing his or her own private narrow bit, such that an overall design, unsought as well as unforeseen by them, is seen to emerge. The point, of course, is that the emergence of the overall design is not left mysteriously unaccounted for, nor, specifically, is it attributed to accident or chance: it is the detailed stages of the invisible-hand process which are meant to supply the mechanism that aggregates the dispersed individual actions into the patterned outcome.

This leads me to the second constraint to be imposed on an invisible-hand process for it to count as adequate. Broadly speaking, it has to do with the ordinariness and normalcy of the stages of the process. By this I mean that, given the circumstances specified at the outset, the story by means of which the invisible-hand process is conveyed has got to sound like a description of the ordinary and normal course of events. It cannot hinge on the extraordinary and the freaky, or on strokes of luck or genius – even if these be logically, physically and humanly *possible*. Mere possibility is too weak a constraint on the type of stories we are after; science fiction-like stories would not yield the kind of explanatory import invisible-hand explanations are expected to afford.[10]

It would be well to comment at this point on the element of *surprise* undeniably involved in invisible-hand explanations and partly responsible for the pleasure and satisfaction derived from the best of them. What is surprising – even startling, sometimes – about invisible-hand explanations is, I suggest, their very existence: the fact, that is, that what one would have thought had to be the product of someone's intentional design can be shown

to be the unsought and unintended product of dispersed individual activity. But once it is realized that it is indeed feasible to provide such an account, there should be no further surprises within the explanation itself. Indeed, among the marks of the good invisible-hand explanations are, I am arguing, the fluency and naturalness of – the very lack of surprising elements in – the description of the process involved.

6. ONE MORE EXAMPLE: SCHELLING'S MODEL OF SEGREGATION

Let me now consider in some detail one more example which will neatly put in relief most of the characterizing features of invisible-hand explanations discussed so far. I refer to Thomas Schelling's [20] thought-up model of residential segregation. His leading conjecture is that "... the interplay of individual choices, where unorganized segregation is concerned, is a complex system with collective results that bear no close relation to individual intent". He is concerned, then, to design abstract models which will actually show how a pattern of rather extreme segregation can be brought about not as a result of central planning or of communal agreement: rather, they will show how such patterns can be the product of the decisions and actions of individuals who, neither intending nor desiring them, only want, e.g., to reside where their own group achieves bare majority status within their immediate neighborhood.

One such model is roughly the following. We are to imagine a line (of houses in a street, chairs at a counter, or the like) along which whites and blacks are distributed, in equal number and in random order. Each of them is supposed to be satisfied in his or her location if their own color group is not in a minority in their immediate neighborhood. Those who are dissatisfied are supposed to move to the nearest point, in either direction, where the condition for satisfaction is met. Specifically, in order to see how this works, we follow Schelling in imagining a line of 70 randomly distributed individuals (35 of each color) and in stipulating that an 'immediate neighborhood' consists of four neighbors (houses) to the right and four to the left. The first stage, then, can be represented by the following diagram, where the 1's stand for blacks and the 0's for whites, and with a dot marking each dissatisfied individual:

0$\dot{1}$000$\dot{1}\dot{1}$01$\dot{0}$0$\dot{1}$1$\dot{0}\dot{0}$111$\dot{0}$11$\dot{0}$11$\dot{0}$0110011001$\dot{0}$0$\dot{1}$10$\dot{1}$0$\dot{1}\dot{0}$0$\dot{1}$11$\dot{0}$1$\dot{1}$00000$\dot{1}\dot{1}\dot{1}$000$\dot{1}$00$\dot{1}$1$\dot{0}$1$\dot{0}$1$\dot{1}\dot{0}$

At this point each dissatisfied individual, starting from the left, is supposed to move to the nearest point where at least half his immediate neighbors are of his own color. (Of course, some of those initially dissatisfied will become satisfied without moving, due to 'desirable' changes in their neighborhood – and vice versa.) So the second stage, which is the result of the first round of moves, is represented by the following:

00000000$\dot{1}$111$\dot{0}$111111111$\dot{0}\dot{0}$00$\dot{1}\dot{1}$000$\dot{1}$01$\dot{0}$111$\dot{0}$111111111000000000000000000111111

(Here there are nine dissatisfied persons, as compared with 26 in the previous stage.)

It turns out – and this is the punch – that the third stage already consists of six stable segregated clusters, with nobody dissatisfied[11]:

000000000111111111111111000000000011111111111111000000000000000000111111

Sifting out of this model the ingredients which are of concern to us, it is evident, first, that the (segregation) pattern to be explained belongs to the social realm, and that it is well-structured. Also, there is little doubt that an explanation of it in terms of intentional design leading to coordinated action is easily forthcoming, whether it be in terms of the execution of a municipal plan, or in terms of the participants' own agreement. Moving on: the account which purports to replace the premeditated-plan one in no way has complete segregation 'in mind'. Its starting point assumptions, which set the invisible-hand mechanism in motion, have to do with the wants (bare local majority) of the individuals involved[12] and with the stipulation of their actions in pursuit of these wants. The process itself consists of well-defined, temporally ordered stages, the last of them being the explanandum phenomenon, the segregation pattern. Furthermore, the stability of this last stage (in terms of the participants' satisfaction in their location) answers the possible further question of why it is that the process actually stops there rather than go on rolling.

The way this model of Schelling's fits into – and thereby helps highlight – the account of invisible-hand explanations which I have been concerned to offer is, then, very neat. But it cannot be ignored that this neatness is bought at a certain price. The price has to do with the fact that the abstractness and quite striking oversimplification of the model rob the

invisible-hand process of some of the required aura of obviousness and naturalness. And yet, it seems to me to be well worth pointing out that although somewhat unrealistic because of the oversimplification, there is nothing weird, outlandish, or abnormal in the assumptions needed in order for the mechanism to be set agoing. This indeed is what in my opinion justifies Schelling's assertion, following his acknowledgment that the model is much too abstract to be thought of as a "motion picture" of what actually takes place in reality, that "it is suggestive of some of the dynamics that could be present in individually motivated segregation".

7. TRUTH AND COGENCY OF INVISIBLE-HAND EXPLANATIONS; THEIR EXPLANATORY AND EXPLICATORY IMPORT

An invisible-hand explanation, like any other explanation, is true when its premises are true; in the present case, when as a matter of historical fact the pattern explained has indeed emerged not through premeditated planning but in a way sufficiently similar to that described by the pertinent invisible-hand process. The merits of such true explanations are obvious. Independently of its truth, however, an invisible-hand explanation may be pronounced 'good' on internal standards: that is, when the degree of structure of the pattern to be explained is high, as are the degree of complexity and sophistication of the invisible-hand process and the degree of conformity of the explanation to the various constraints laid down so far. Let us agree to refer to such 'internally good' explanations as *cogent*.

Now invisible-hand explanations that are both cogent and true are too meritorious to give rise to any queries. One can only wish there were more of them. But queries do arise with regard to invisible-hand explanations that are cogent while (factually) false, as well as with regard to those cogent ones whose truth is undetermined (or maybe undeterminable): what, if any, is their worth?

Let me begin with the cogent invisible-hand explanation which is actually known – or eventually proven – to be false. Suppose, for example, that newly found archives in Venice establish it beyond doubt that, with the foundation of the Banco della Piazza di Rialto in the 16th century, Doge Nicolo da Ponte had been sufficiently shrewd and enlightened to recognize that a reserve ratio of a hundred per cent was not necessary and, furthermore,

that if it were to be less, then the bank could also fulfil the useful function of adding to the money supply in the state; and so he then and there thought out – and put into effect – the whole commercial banking system of Venice, reserve ratio, monetary expansion and all.

What then of our goldsmiths' story? Does the fact that the Doge beat the invisible-hand explanation to it rob the invisible-hand explanation of its explanatory import? I suggest that the answer is No, and that the argument for this answer goes beyond the mere 'feeling' we may have that the account of how something could have arisen without anyone devising it is 'interesting' or 'illuminating' in its own right.

The argument I should like to put forward is that 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*. Not every product of design, especially if we are dealing with a complex social pattern or institution, is successful and lasting. The ascertainment that there is (was) a designing agent, therefore, even when conjoined with the ability to identify him and to spell out his rationale, does not take us very far towards illuminating the nature of its success and stability. The availability, on the other hand, 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 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.

There is however another possibility. The success and durability of the intentionally designed social pattern or institution under study may in some cases be attributable to its performing a useful function. In such cases, then, the story of how it could have arisen without anyone devising it will not only not be the true account of its emergence, but it will also be of secondary importance (at best) as an account of its durability. I shall discuss this possibility below, in the broader setting of the issue of functional-evolutionary explanations.

Moving on now to cogent invisible-hand explanations whose truth is undetermined we note, first, that all that was said above concerning the explanatory import of cogent-yet-false invisible-hand explanations a fortiori holds here too. That is, even if we are unable, on the basis of available

(historical) evidence, to select as true one of the two competing emergence hypotheses – the intentionally devised vs. the invisible-handedly evolved, the invisible-hand explanation will not be devoid of explanatory import as far as the continued existence of the pattern in question is concerned.

But here a further, intriguing, question may be posed. When a cogent invisible-hand explanation is given to a phenomenon heretofore 'naturally' supposed – but not proved – to have come about through intentional design, can anything at all be said about the probability that *it* is the correct account? Can the fact that an invisible-hand explanation is hit upon count as evidence in support of its own truth?

I have argued earlier (Section 4) that the more structured the phenomenon to be explained, the more easily forthcoming would an intentional-design explanation of it be, and hence the more plausible will this explanation appear. I would now like to suggest, further, that it will also be judged to be highly *probable* that some intentional-design explanation is the factually correct explanation – so long as the only alternative conceived is that the phenomenon concerned emerged through some unaccounted-for 'accident'. The reason, of course, is that the relative frequency of highly structured social patterns that come about through planning is higher than those whose origins can be attributed to mere accident. However, once an invisible-hand explanation of that same phenomenon comes upon the stage, and supposing it to be a cogent one, the comparison changes: the intentional-design account will no longer be weighted for probability against a non-account of accident, but rather against an account in terms of an explicit and detailed invisible-hand mechanism. My conjecture here is that even though its own probability cannot be determined a priori, the mere availability of a cogent invisible-hand explanation does indeed undercut the probability of the intentional-design account it purports to displace. I am, however, in possession of no further arguments capable of turning this conjecture into a proper claim. And so, unhappily, I leave the matter at that.[13]

The discussion in this section focused so far on the *explanatory* import of invisible-hand explanations. A further issue is that of their *explicatory* import. Regardless of whether or not an invisible-hand explanation is the correct account of how the pattern in question actually emerged, I contend that it is of value as an (at least partial) explication of the associated concept. By this I mean, first of all, that an account of how something could have emerged rather than the tracing down of its actual origins is generally viewed

as a rational reconstruction of it, and as such it is taken to perform an explicatory task (see Carnap [3], pp. 576–7).

Now with regard to our particular object of concern – cogent invisible-hand explanations – there are, I believe, two further considerations that enhance this explicatory value. First, the sense of 'could' in the expression 'how a pattern could have emerged' is here severely restricted. As was argued earlier, it is required not only that it be logically, physically and humanly possible for the pattern in question to have arisen in the specified way, but that this specified way actually conforms with what can reasonably be expected to be the normal course of events. Second, there is the element of reduction: the pattern in question is shown by the invisible-hand explanation to arise through a process involving the actions of individuals who do not have it in mind. Thus, there should in principle be no reference to the explanandum phenomenon within the spelled out explanation; the linguistic expression which stands for it need neither be used nor mentioned in the course of the explanation.

It seems to me that these two considerations taken together go a long way toward justifying our intuitive feeling that 'good' invisible-hand explanations yield substantial understanding of, and provide considerable insight into, the nature of the social patterns and institutions explained, and that as such they are of remarkable explicatory worth. There is even a sense in which the fact that a cogent invisible-hand explanation proves false is felt to be peculiarly irrelevant: the fact that someone was actually smart and quick enough to have intentionally brought about the pattern in question is felt, I think, to shed but little light on its nature – indeed is felt to be almost accidental.[14]

Since the next section will introduce a different approach to invisible-hand explanations it would be well, I think, to pause at this point and take stock. Put schematically, the distinguishing features of invisible-hand explanations that emerge from the entire discussion so far are the following.

The domain of explanation: The social domain.

The explained phenomena: Well-structured social patterns that result from human action and that look like the product of intentional design.

The nature of the explanation: A species within the genus of genetic explanations, it consists of a description of a special kind of ('invisible-hand') process, and is addressed to the question of the generation of the explained phenomenon.

The mode of explanation: Displacing the intentional-design account that 'naturally suggests itself' with an account that specifies the working of a mechanism that aggregates the dispersed actions of individuals into the overall pattern (the explanandum phenomenon), subject to the assumption that the individuals concerned neither foresee this resultant of their actions nor intend to bring it about.

Idiosyncrasies of the explanation: Its having both explanatory and explicatory import; its lending itself to be characterized as 'cogent' in addition to the usual categories of truth and falsity; its being a matter of degree (varying with the extent to which the explanandum phenomenon is structured and with the complexity of the invisible-hand process); the element of surprise attaching to it.

I prefer to regard all these as distinguishing features of invisible-hand explanations rather than as necessary and/or sufficient conditions – or as adequacy conditions, for that matter – for the following reason. I believe that the class of invisible-hand explanations is not at all sharply delineated (indeed, I take it that the picture drawn so far attests to that). The best and highest that one may aim at, therefore, is to characterize what one takes to be the *core* of this class, namely its undisputedly paradigmatic members. This is, of course, a matter of exercising judgment, and as such there is always the possibility of being persuaded this way or that. Counterexamples and problematic examples may be found and cited. They will affect the picture here drawn, however, only to the extent that they can be shown to relate to its core rather than to its fringes.[15]

8. A DIFFERENT APPROACH: THE FUNCTIONAL-EVOLUTIONARY MOLD

There exists an altogether different approach to invisible-hand explanations, one that has to do with the nexus of functional-evolutionary explanations. In a rough and preliminary formulation, it views invisible-hand explanations as the counterpart, within the social domain, of the biological-evolutionary explanations within the domain of living organisms. In order to be able to evaluate this approach, as well as to compare it later to the one offered so far, I propose to take some space to draw the broader picture that I take to be its proper setting.

There is an abundance of recent literature on functional analysis. This is not the proper place to survey it.[16] I shall, however, venture to put forward what, on the basis of this literature, I believe to be a formulation that conveys the gist of the explanatory work a function-ascribing statement is supposed to do. It is the following:

> The function of x in a system s is ϕ

means

> The system s has capacity/goal ψ and the ϕing of x is an essential element in the explanation of ψ.

(The idea behind this schematic formulation is that the function of something in a system is to be construed as its contribution to the way the system operates, or to the way it attains its goal.)

There are two major areas where function-ascribing statements frequently occur: in the explanation of parts of artifacts, and in the explanation of organs of living organisms. Thus, a typical functional analysis involving artifacts will be that of the gnomon: the item, x, is the gnomon, the system, s, is the sundial, the function, ϕ, of x is the casting of a shadow on the dial beneath, and the capacity/goal, ψ, of s is to show the correct time of day. And, analogously, a typical functional analysis involving organisms will be that of the kidneys: the item, x, is the kidneys, the system, s, is the living body, the function, ϕ, of x is the elimination of wastes from the blood, and the capacity/goal, ψ, of s is its survival and well-being (or something of the sort).

It is worth noting, in passing, that while in the case of artifacts the s and the ψ – the system and its capacity/goal – vary from one analysis to the other, in the case of organisms they are virtually constant: it is usually relative to the living organism as a whole[17] and to its capacity/goal of survival that a function is attributed to some component thereof. This point will be referred to in the sequal.

Now functional analysis, however it may be construed, has as a rule been expected to be closely related to the causal story pertaining to the question of *why x is there*. A good deal of controversy and confusion, however, surround this connection.[18] As I see the matter, in the standard and archetypal cases where functional analysis is used, the interpretative formula offered above is indeed closely related to the causal-genetic story of why the

functional item is 'there'. The close relation can be put thus: the functional analysis of an item yields an account of its presence through the mediation of certain specific though implicit assumptions which are – again, in the paradigmatic cases – more or less safely taken for granted.

Let me elaborate. In the case of artifacts, the mediating assumptions have to do with the intentions and beliefs (reasons, if you will) of the designer. Thus, having subjected the gnomon to functional analysis along the lines suggested above, we may arrive at its causal-genetic account by augmenting it with such assumptions as: the designer intended to produce a sundial capable of indicating the correct time of day, on the basis of his past experience and observations he believed that if the gnomon were to be installed in the particular place at the particular angle it would cast a shadow in the desired way, and, finally, that this intention and this belief led him to put it there.

In the case of the organism the mediating assumptions are drawn from the theory of evolution. The causal link here, however, is subtle and apt to be misunderstood. For one thing, the role played by natural selection is *not* analogous to that played by the designer's reasons in the artifacts case: natural selection does not 'lead' to the 'installment' of anything. The ultimate factor causally relevant to the presence of organs is that of random mutation that operates on the organism's genetic plan.[19] Thus, having subjected the kidneys to functional analysis along the lines suggested above, we may arrive at, or maybe only approximate, their causal-genetic account by augmenting it with something like the following: in vertebrates whose genetic plan specifies, due to successive mutations, a kidney-like organ, wastes are efficiently removed. Hence, by the very performance of their function, the kidneys help vertebrates incorporating them to 'succeed', i.e. to survive and be selected for, and thereby they contribute to their own *continued presence* in them.

Having said all this, I would like to emphasize that this relation between the functional analysis of an item and a causal-genetic account of its presence, although often close (in the sense just spelled out), is by no means necessary. A functional analysis of something cannot, in and of itself, be construed as an answer to the question of why this something is 'there'. In many non-standard cases the analysis of the function-ascribing statement provides no clue to, indeed has nothing whatsoever to do with, an account of "how it came to be what it is and where it is".[20]

I have in the foregoing remarks spoken of functional analysis as used in the two major domains of artifacts on the one hand and organic nature on the

other. It is time now to consider how it fares in the social domain which, as indicated in the opening remarks (Section 1), can be viewed as in a sense interposed between these two.

The scheme of functional analysis presented above can in my view adequately handle cases in the social domain too.[21] Thus (to take a familiar example), it is pointed out by sociologists and anthropologists that the ostensibly superstitious and idle rain-making ceremonies of the Hopi may nevertheless be taken to fulfil an important function – to be sure, a *latent* one – that of "reinforcing the group identity by providing a periodic occasion on which the scattered members of a group assemble to engage in a common activity" (Merton [14], pp. 118–9). So here the item, x, to be functionally analysed is the rain ceremonials, the (social) system, s, is that of the Hopi, the function, ϕ, attributed to x is the enhancement of group unity and cohesion, and the capacity/goal, ψ, of s is its equilibrial survival in a state of "social health" (the phrase is Radcliffe-Brown's [18], p. 183).

Before going on let me allude to a point made earlier (p. 279) and remark that like in the case of parts of living organisms (– and unlike the case of parts of artifacts –) so also in the case of a social pattern or institution the system, s, and its capacity/goal, ψ, are virtually constant: it is relative to the social unit as a whole and to its capacity/goal of survival that a social institution is functionally analysed.

Now what, if any, is the relation between the functional analysis of a social pattern or institution and a causal-genetic account of its presence? It seems to me that the discussion up to now (and in particular the remark of the preceding paragraph) fairly clearly suggests that the case of functional analysis as applied to the social domain is closer in nature to the cases of its application to organisms than to artifacts. And if so, it is to be expected that the type of connection between this analysis and the causal-genetic account will be affined to that which was shown to obtain within the context of organic nature rather than to that within the context of artifacts. That is to say, the mediating assumptions required for this connection may plausibly be expected to be *evolutionary* in kind.

The intimate relationships – indeed the quite conscious cross-fertilization – between the theories of biological and social evolution hardly require documentation here;[22] besides, their elaboration right now would be beside the point. Coming to the point, my contention is the following: according to 'the other approach' to invisible-hand explanations, as I

understand and interpret it, it is the functional analysis of the pertinent social institution, conjoined with the concomitant evolutionary apparatus presumed to supply the missing causal link, that constitute – together – the invisible-hand explanation of that institution.

To wit: when a social pattern or institution is to be explained within the framework of this approach, then first of all its function within the relevant social unit has to be ascertained. That is, an effort is to be made to find out its contribution (if any) to the equilibrial and frictionless survival of the society in question. Once this is successfully established, the phenomenon under study is assumed all but explained, the (implicit) filling in being that by performing its function even its faint beginnings – whatever their origins – are reinforced and selected for; consequently this institution is better capable of helping the social unit incorporating it to 'succeed', and this 'success' of the social unit, in turn, accounts for the institution's own perpetuation in it. Now this process of *selection* is supposed to be a non-man-made one: it is visualized as a large scale evolutionary mechanism that as it were scans the inventory of social patterns and institutions at any given period of time and screens through to the next those of them that are best adapted to their (respective) roles. Whence the designation of 'invisible-hand' to this mode of explanation.

Notable among the authors to whom this conception of invisible-hand explanations may be attributed is F. A. Hayek; he is also the most explicit in what he says about this matter. According to him, social institutions "did develop in a particular way because the coordination of the actions of the parts which they secured proved more effective than the alternative institutions with which they had competed and which they had displaced". And he goes on to state: "The theory of evolution of traditions and habits which made the formation of spontaneous orders possible stands therefore in a close relation to the theory of evolution of the particular kinds of spontaneous orders which we call organisms." ([7], p. 101)

9. THE TWO MOLDS OF INVISIBLE-HAND EXPLANATION: A COMPARISON

We started out with a quest for a coherent notion of invisible-hand explanations. It now seems that we may have ended up with more than we

had bargained for. For we now seem to be in the possession of two – let us call them *molds* – into which an invisible-hand explanation can apparently be cast: for convenience of reference I propose to label the first the *aggregate* mold and the second the *functional-evolutionary* mold. How are they related? – Do they converge upon, complement, or perhaps compete with each other?

I would like to begin with a delineation of the ground common to both of these two molds of explanation; this should also reveal why they both claim to the title of 'invisible-hand'. It seems to me that the motivation behind these two molds of explanation is one and the same: it is to provide mature explanations to phenomena from the realm of human action. The sense in which these explanations are supposed to be *mature* is that they are purportedly liberated from the grip of the formative – yet in certain respects primitive – picture according to which to explain is to point to a designer (conspirator), or creator. Thus, an explanation which fits either of these molds is supposed to account for a social pattern or institution as a spontaneously formed order, to account for its existence subject to the assumption of no outside intervention by any designing agency.

'To account for something's existence', however, can be construed in various ways. And, indeed, it is precisely when the question of what it is that we want explained is pushed, that the distinct contours of the two molds begin to take shape. What I shall now proceed to argue is that the basic questions to which the two molds of explanation under consideration are respectively addressed are different. Furthermore, what may be termed the *background interests* underlying these basic questions are different too. The spelling out of these differences will facilitate the assessment of the points of divergence between the two types of explanation themselves, and also between the types of social phenomena they come to explain.

The characterization of the first – the aggregate – mold of invisible-hand explanations has taken up the bulk of this study. This characterization has of course provided us with an answer to the question of what it is that we want explained within the framework of this mold: given a certain social pattern or institution, it is first and foremost asked *how did it – or how could it have – come about*? This question, however, is laden. Unpacked, it turns, as we know, into a quest for one particular mode of emergence rather than another: it is a quest for an account in terms of an invisible-hand process

rather than for an account, which is assumed from the outset to be readily available, in terms of intentional design.

The basic question with which an explanation of the functional-evolutionary mold is concerned, on the other hand, is this: given that a certain social pattern or institution exists, *why is it in existence*? This question, too, has to be unfolded in accordance with the interests underlying it. It then becomes: Why is the social item under study existent rather than non-existent, or, how come it has not by now been eliminated, or, again, why does *it* exist rather than some alternative ('functionally equivalent') social arrangement. It should be noted and emphasized that an explanation of this type involves no commitment as to how the scrutinized pattern actually originated.[23] For all that it tells us the pattern in question could have come into being as a result of intentional design and careful execution, or, for that matter, it may have originated (somehow) through people's "stumbling upon establishments, which are indeed the result of human action, but not the execution of any human design".[24] An explanation that belongs within this mold decidedly leaves the matter of the origins of the explained phenomena as unaccounted for as does the notion of spontaneous and random mutation within the theory of biological evolution.

It is evident, then, that we are dealing with two quite disparate sets of questions. It follows that the molds of invisible-hand explanation corresponding to them constitute two quite disparate undertakings: the first is concerned with providing *a chronicle of* (a particular mode of) *emergence*, the second with establishing *raisons d'être*.

Further differences pertain to the social patterns or institutions eligible for explanation in either of the two molds. The explanatory onus of an invisible-hand explanation of the aggregate mold is, as was shown, on the invisible-hand process that accounts for the emergence of the item in question. The point to be stressed is that there are no prerequisites attached to the social item itself, except that it be *structured* in some interesting sense.[25] Matters are different though with an invisible-hand explanation of the functional-evolutionary mold. There it is a prerequisite that the pattern or institution to be explained have a *function*, that it perform a useful service to the social unit incorporating it. To be sure, once we allow ourselves to get immersed in the so-called functionalist school in sociology and anthropology and to adopt its viewpoint we ought presumably to grant that rather than

requiring that the explanandum phenomenon have a function, it is indeed *presupposed* that it has one (or else it would not have arisen in the first place, or would have been eliminated in due course). But then there is of course a host of further presuppositions that this framework will commit us to: that any existing social pattern or institution, besides being 'functional', is also 'optimal' (as a 'solution') in some sense, that organic social units are self-regulating systems capable of and aiming at survival, and more. In other words, while the exposition of the aggregate mold of invisible-hand explanations constitutes a methodological undertaking dealing mainly with the formal aspects of an explanation scheme, the exposition of the functional-evolutionary mold is, in my opinion, inherently bound up with certain assumptions of substance – if not downright ideology – concerning the nature of the explained phenomena and of society in general.

Having dwelt upon the points on which the two compared types of explanation diverge, let me now, finally, indicate where they may converge – or, rather, how they may be superimposed. The point of junction is the following. Suppose a certain existent social pattern or institution is up for explanation. The attempt may first be made to subject it to the first, aggregate, mold so that – if successful – an invisible-hand account of how it (could have) emerged is obtained. It may then be pointed out that this item is functional, that it contributes to the well-being and proper working order of the society incorporating it. So, it may now be subjected to the second, functional-evolutionary, mold, yielding an invisible-hand account of its durability and prevalence (if indeed it is prevalent across societies and cultures).

Thus, suppose that the emergence of the monetary system is shown – along the lines suggested in Section 2 – to be due to an invisible-hand process rather than to someone's intentional design. Suppose, further, that the (quite plausible) case is made that this 'institution' of money performs a role useful to the smooth and efficient functioning of the society in which it has emerged. At this point the evolutionary apparatus may be applied to it and as a result the continued existence of the money institution, indeed its spread, will be accounted for. That is, given that the social pattern or institution to which the first mold of invisible-hand explanations has been successfully applied is functional, the second mold may also be applied to it. This superimposition of the two explanatory molds may be expected to yield a

rather satisfying answer to the question concerning the existence of this institution, since it is capable of yielding an account of its emergence and inherent stability as well as its continued existence and prevalence.

Now it is possible that many – or most, or even all – of the social patterns whose emergence is invisible-handedly explained within the framework of the first mold will be found to be 'functional' and hence amenable to the second mold as well. For myself, I prefer to withhold judgment on that; indeed, I take Schelling's model of segregation (Section 6) to serve as a warning against too easy an acceptance of a general statement on the matter. Be that as it may, however, I would like to close by emphasizing that if I am anywhere near the mark about there being a distinction between the two molds of invisible-hand explanations, it seems to me important to keep them distinct and to conceptually isolate them from each other.

The Hebrew University of Jerusalem

NOTES

*This essay was written in memory of Yehoshua Bar-Hillel.

I owe a deep debt of gratitude to Avishai Margalit for indefatigable help in advice, criticism, and ever-attentive listening. I also want to thank Robert Nozick who helped shape my ideas both through the pages of his book that deal with this subject and in conversations while this paper was being written; he also drew my attention to some of the items in the bibliography. Sidney Morgenbesser read the final version; his insightful comments and reservations led to the insertion of (– and are, I hope, adequately expressed by –) notes 1, 10 and 15. I am truly grateful to him.

[1] There is a sense, of course, in which the notion of unintended consequences of human action was not only known since ancient times, but was in fact pivotal to classical tragedy. Thus, Oedipus' action, done with the intention of hitting the man in the chariot, resulted in the undesigned killing of his father. It is not, however, of such actions, which may be thought of as accidents, that the 'third realm' is supposed to consist. More on this and related issues in Section 3 below.

[2] For an account of the history of writings on this 'middle realm', as well as for a vigorous statement of its importance, see Hayek [6], Ch. 2, and [7].

[3] For the niceties of the full story see Samuelson [19], Ch. 15.

[4] I would like to refer the reader to a recent article by Jones ([12]), in which he offers precise empirical models – albeit in a very specialized context and subject to rather severe restrictions on the 'trading environment' – that show how media of exchange actually emerge through the unconcerted market behavior of individuals. Noteworthy among the points made in this article are the following: (a) A stable exchange pattern is obtainable not only by full monetarization but also by 'some mixture of direct barter

and use of a common good as medium of exchange'; (b) There are reasons to expect that the first commodity money to emerge in a barter economy would in fact be a very common good rather than a scarce one (like rice in Japan, beaver pelts in Canada, and more).

[5] It should be remarked, though, that these days any discussion of the central nervous system is likely to be itself dominated by the influence of yet another model, that of computers, electronic control, automatic command centers, and the like.

[6] For Robert Boyle the model for the 'machine of the world' was the famous cathedral clock at Strasbourg, in which "the several pieces making up that curious Engine are so fram'd and adapted, and are put into such motion, as though the numerous wheels and other parts of it knew and were concerned to do its Duty". (*The Usefulness of Experimental Natural Philosophy*, 1663.) That is, not only does the existence of the clock attest to its designer, but – according to Boyle – the supreme coordination and adaptation of its parts (almost) attest to their being conscious and intentional agents.

[7] The well-known Smith passage concerning the invisible hand occurs in *The Wealth of Nations*, IV.II.9. Somewhat less known is the earlier occurrence of this concept in his *Moral Sentiments* (the first edition of which was published in 1759), in a remarkable passage dealing with the distribution of means to happiness that is brought about when the landlords who, seeking to ever enhance their wealth and greatness but whose "capacity of [the] stomach bears no proportion to the immensity of [their] desires", are forced to divide with the poor what they produce above their own needs:

> They [the rich] are led by an invisible hand to make nearly the same distribution of the necessaries of life, which would have been made, had the earth been divided into equal portions among all its inhabitants, and thus without intending it, without knowing it, advance the interest of the society, and afford means to the multiplication of the species. (*The Theory of Moral Sentiments*, IV.1.10. The Glasgow Edition (1976): Vol. I, pp. 184–5.)

The very first use of the expression 'invisible hand', however, appears to have been made by Smith in his *History of Astronomy*, III.2. But there it seems to have had theological connotations having to do with supernatural intervention. See A. L. Macfie, 'The Invisible Hand of Jupiter', *Journal of the History of Ideas* **32** (1971), 595–599. See also T. D. Campbell, *Adam Smith's Science of Morals*, George Allen and Unwin Ltd., London, 1971, esp. pp. 60–62, 71–73, 117(n). I am grateful to Allen Silver for having drawn my attention to these passages and references.

[8] The emphasis in Hempel's discussion, beautifully illustrated by two *historical*-genetic examples, is on the combination in these explanations of "a certain measure of nomological interconnecting with more or less large amounts of straight description" ([11], p. 449); this holds here too. As is being shown in the text, however, I take it that the invisible-hand processes are subject to more specific, as well as more severe, constraints than are the sequences in ordinary historical-genetic explanations (except possibly for the requirement of truth – see Section 7 below).

[9] Nozick notes ([16], p. 352, note 7) that there will be cases where some of the participating individuals foresee the ultimate outcome and intend to bring it about while others do not, and suggests therefore that the notion of invisible-hand processes might be refined to admit of degrees. Compare this to my quite different notion of invisible-hand explanations' being a matter of degree (p. 267 above).

[10] Will the stipulation that the participating individuals are *rational* take care of this normalcy condition? It would seem so: the rationality of the agents' actions, relative to their beliefs and utilities as determined by the circumstances prevailing at the initial stage – as well as by the possibly changed circumstances of subsequent stages – would indeed go a long way toward insuring that the stages of the sequence not be startingly strange or deviant. There is, however, one level at which the requirement of rationality does fall short of the requirement of normalcy, namely the level of the beliefs and utilities themselves. That is, in addition to the requirement that the agents act rationally relative to their beliefs and utilities, we would not want to allow the invisible-hand explanation to hinge in an essential way on crazy or freakish beliefs and utilities, relative to the prevailing (and possibly changing) circumstances – unless, of course, there is *independent* anthropological evidence that such indeed were the beliefs and utilities of the people involved. It follows that the normalcy constraint, as sketchily presented in the text, may in fact be broken up into two componenets: the normalcy (within some vague range) of the agents' beliefs and utilities, given the circumstances, and their acting rationally, given these beliefs and utilities.

[11] As Schelling points out, this is not the only stable 'solution' that satisfies everyone's minimum requirements. Thus, alternating 1's and 0's, or alternating pairs of 1's and 0's, will also do – as would clusters of five. But *this* is what comes out under the 'circumstances' of the particular example, i.e. with the particular random distribution of 1's and 0's we began with. (A question that seems to me worthy of consideration is this: Is there a randomly distributed series of 1's and 0's such that, were *it* to be the first stage in the process, the very same assumptions of Schelling's model would lead to its turning into a series of alternating 1's and 0's? Or is it the case that in order for this desegregated outcome to be brought about an intervening hand is inevitably called for?)

[12] It is of importance to note that each participant's desire to achieve bare majority status, far from being a disguised desire for total segregation, is in fact quite compatible with the further desire to reside in a heterogenous rather than in a completely homogenous neighborhood. However, if indeed we attribute to our participants this further wish, then, as Schelling points out, it turns out that almost half of them remain dissatisfied on this score, since thirty out of the seventy actually end up in a neighborhood of their own color exclusively.

[13] Matters are somewhat reversed with F. A Hayek, who vehemently denounces "that factually untrue anthropomorphic interpretation of grown institutions as the product of design" ([7], p. 102): for him, it seems, every social institution is presumed to have evolved invisible-handedly, unless proven otherwise. Thus, in the context of justifying some of the criticisms levelled against Adam Smith's account of the market mechanism he emphasizes that "his [Smith's] implied assumption, however, that the extensive division of labour of a complex society from which we all profited could only have been brought about by spontaneous ordering forces and not by design was largely justified" ([7], p. 100). This position of Hayek's, however, should be appraised in the light of the second, functional-evolutionary, conception of invisible-hand explanations, to be introduced and discussed in Sections 8 and 9 below.

[14] While mentioning this curious possibility, that someone's intelligent design might come to be seen, in the light of an alternative invisible-hand account, as in a sense accidental, let me allude to Hegel's notion of the *Cunning of Reason* ('List der Vernunft'). His idea seems to be that while the leader appears to be concerned to do 'his

own thing' – to act on his own private motives – his actions, put in the wider historical perspective, are seen to fit into an overall (historical) design. Thus, the 'true' leader appears to be led by means of the 'cunning of reason' to do that which inevitably had to be done. Consider: "Thus he [Caesar] was motivated not only by his own private interest, but acted intrinsically to bring to pass that which the times required." ([8], pp. 39–40.) My interpretative reconstruction of this, suggested by the present approach to invisible-hand explanations, would be the following: A leader may (sometimes) bring about, in an initiated and premeditated way, some social pattern or institution that might otherwise have emerged – later – through an invisible-hand process. In that sense it may (perhaps) be said that, while manifesting in his deeds the cunning of reason, the leader is an 'instrument in the hands of history'. On this view see also Hegel [9], Section 199 (p. 129), and Taylor's comment upon this section ([21], p. 433) in which he points out the connection between Hegel's thesis of the cunning of reason and Adam Smith's thesis of the invisible hand.

[15] Having offered this interim recapitulation I would like to suggest several areas where the scope of the present analysis may possibly be broadened, thereby conferring the status of invisible-hand explanations upon cases that deviate from the core class in the systematic ways to be indicated. (a) The case can be made that an invisible-hand explanation may be offered to account not just for the *generation* but also for the *degeneration*, or disintegration, of a well-structured social pattern. A strict analogy holds between these two types of cases. (b) Given that the explanation I have been talking about consists basically of an ordered couple, the first component of which being the social pattern to be explained and the second the *expectation* that it be explained in terms of intentional design, it may plausibly be argued that the latter could be relaxed, or even ultimately abandoned. That is, it seems quite plausible to suggest that an invisible-hand explanation may be offered where there is *no* expectation at all as to what kind of explanation is called for, where one simply does not know. Moreover, if acquaintance with some (good) invisible-hand explanations that have been offered in the past and have proved successful can be assumed, one may indeed start out with the expectation that the particular social institution that is up for explanation right now would (should) be invisible-handedly explained. This shift in the expectation component, however, is thus clearly not altogether causally independent of the (known) success of some invisible-hand explanations that do belong to the core class as presented in the text. (From among the distinguishing features of invisible-hand explanations the one that will be undermined by the relaxation of the expectation condition is, obviously, the surprise factor.) (c) The requirement of methodological individualism may be relaxed to allow units larger than the single individual, such as households or firms, to be the deliberating, deciding, and ultimately acting participants.

[16] Some of the central articles to have appeared recently on this subject are those of Boorse ([2]), Cummins ([4]), Wimsatt ([22]), and Wright ([23]). I want to record here a special debt to Cummins's stimulating paper; some of the examples discussed in the sequel are drawn from it.

[17] Possible variations here occur when the system s is taken to be a whole species, or even the entire eco-system. (See Boorse [2] and further references there.)

[18] In addition to the articles referred to in note 16 above, all of which deal with this issue, one should of course mention Hempel's essay ([10]) which is the *locus classicus* of all modern discussions of functional analysis (and, at least according to Cummins, also

the main source of confusion regarding this question of the connection between the functional analysis of an item and a causal account of its presence).

[19] Cummins [4]: "We could therefore think of natural selection as reacting on the *set* of plans generated by mutation by weeding out the bad plans; natural selection cannot alter a plan but it can trim the set."

[20] The quoted phrase is Radcliffe-Brown's ([18], p. 186). For good examples of cases of functional analysis in non-standard contexts see Boorse [2], especially those that have to do with what he calls "accidental functions". (E.g.: the functional analysis of the bible in the soldier's breast pocket, which (happened to) function as a bullet shield, surely does not bear on its presence there.)

[21] One of the main points of criticism against Boorse's article ([2]) is, in my opinion, the striking absence of the social realm from it. Moreover, one of the central assumptions underlying his proposed formula of functional analysis is that the system relative to which the analysis is offered is a 'goal directed' one – and I would seriously question the applicability of this assumption (without ideological presuppositions) to the social realm.

[22] As is well known, it was from the rather well-entrenched views – predominantly due to the 18th-century Scottish school – concerning social and cultural evolution that Darwin and his contemporaries in fact derived the suggestion for their biological theories. Towards the end of the 19th century the biological notions were reimported into (as well as abused by) the social sciences, mainly in the form of the so-called Social Darwinism. (On this topic see, e.g., Hayek [6], p. 59ff.)

[23] Consider, e.g., Merton's careful wording, when discussing the (latent) functions of the political machine: "*Whatever its specific historical origins*, the political machine *persists* as an apparatus for satisfying otherwise unfulfilled needs of diverse groups in the population." ([14], p. 127, emphasis mine.) Radcliffe-Brown, too, distinguishes between accounting for the 'history' of a social institution and accounting for its functioning (see, e.g. [18], p. 186). On the other hand I believe that for both Malinowski (e.g. [13], p. 139) and Hayek (e.g. [7], p. 101) the distinction is obscure, and that for both of them to explain the manner of functioning of a social institution is at the same time to answer the question of its origin or formation.

[24] The phrase – it's about time that it be acknowledged – is Adam Ferguson's ([5], p. 187). Let me mention in this connection Alchian's [1] attempt to apply evolutionary ideas to economic theory. The greater part of his argument is compressed in the following: "The economic counterparts of genetic heredity, *mutation* and natural selection are imitation, *innovation* and positive profits." (Emphasis mine.) He also points out that innovation may be achieved not only premeditatively but unpremeditatively as well – through imperfect imitation, for instance.

[25] I venture, hesitantly, to suggest that whether or not the social pattern can be easily referred to by means of a conventionally codified designation ('medium of exchange', 'minimal state', 'segregation', 'free market', etc.) has something to do with its being 'structured' in the sense I am after.

BIBLIOGRAPHY

[1] Alchian, A. A., 'Uncertainty, Evolution, and Economic Theory,' *Journal of Political Economy* 58 (1950), 211–221.

[2] Boorse, C., 'Wright on Functions,' *Philosophical Review* 85 (1976), 70–86.
[3] Carnap, R., *Logical Foundations of Probability*, Routledge and Kegan Paul, London (1950), 1962.
[4] Cummins, R., 'Functional Analysis,' *Journal of Philosophy* 72 (1975), 741–765.
[5] Ferguson, A., *An Essay on the History of Civil Society*, London, 1767.
[6] Hayek, F. A., *The Constitution of Liberty*, Gateway Edition, Chicago (1960), 1972.
[7] Hayek, F. A., 'The Results of Human Action But Not of Human Design', in his own *Studies in Philosophy, Politics and Economics*, Routledge and Kegan Paul, London, 1967, pp. 96–105.
[8] Hegel, G. W. F., *Reason in History*, The Library of Liberal Arts (Bobbs-Merrill), 1953.
[9] *Hegel's Philosophy of Right*. Translated with notes by T. M. Knox. Oxford University Press, Oxford, (1952), 1967.
[10] Hempel, C. G., 'The Logic of Functional Analysis'. In [11].
[11] Hempel, C. G., 'Aspects of Scientific Explanation', in his own *Aspects of Scientific Explanation*. The Free Press, New York, 1965.
[12] Jones, R. A., 'The Origin and Development of Media of Exchange', *Journal of Political Economy* 84 (1976), 757–775.
[13] Malinowski, B., 'Anthropology', in *Encyclopaedia Britannica*, 1st Supplementary Volume, The Encyclopaedia Britannica, Inc., London and New York, 1926.
[14] Merton, R. K., *Social Theory and Social Structure*, Enlarged edition, The Free Press, New York, 1968.
[15] Mises, L. V., *The Theory of Money and Credit*, 2nd edition, Yale University Press, New Haven, 1953.
[16] Nozick, R., *Anarchy, State, and Utopia*, Blackwell, Oxford (Basic Books, New York), 1974.
[17] Popper, K. R., *The Open Society and Its Enemies*, Vol. 2, Routledge and Kegan Paul, London (1945), 1966.
[18] Radcliffe-Brown, A. R., *Structure and Function in Primitive Society*, Cohen & West, London, 1952.
[19] Samuelson, P. A., *Economics*, McGraw-Hill, 4th edition, New York, 1958.
[20] Schelling, T. C., 'Models of Segregation', *American Economic Review* 59 (1969), 488–493.
[21] Taylor, C., *Hegel*, Cambridge University Press, Cambridge, 1975.
[22] Wimsatt, W. C., 'Teleology and the Logical Structure of Function Statements', *Studies in History and Philosophy of Science* 3 (1972), 1–80.
[23] Wright, L., 'Functions', *Philosophical Review* 82 (1973), 139–168.