

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

INVISIBLE-HAND EXPLANATIONS RECONSIDERED

Randall R. Curren



SOCIAL SCIENCE WORKING PAPER 621

February 1987

INVISIBLE-HAND EXPLANATIONS RECONSIDERED

Abstract

Edna Ullmann-Margalit introduced the notion of an invisible-hand explanation (I-H explanation) to the philosophical literature in 1978, and made a distinction between "aggregate" and "functional-evolutionary" (F-E) forms of I-H explanations. The present paper produces a substantially refined analysis of the forms and functions of I-H explanations. Sections (1) and (2) introduce the ideas of I-H and aggregate I-H explanation, respectively. Section (3) argues that no one form of explanation can serve the explanatory functions Ullmann-Margalit attributes to aggregate explanations, and divides those explanatory functions between genetic and "systematic-dispositional" explanations. Section (4) identifies difficulties with the idea of F-E explanation in the social realm, and shows that any I-H explanations fitting the F-E mold would constitute simply a special class of "aggregate" explanation.

INVISIBLE-HAND EXPLANATIONS RECONSIDERED*

The identification of a distinguishable but hitherto undistinguished form of scientific explanation is a rare and stimulating event in philosophy, its value consisting in the fact that it may be expected to provide the philosopher with not only a new object of study, but also a liberated conception of the possibilities for explanation in domains other than those in which the new form is first discovered. How much work such a discovery creates for those interested in the structure of explanation varies inversely, of course, with the exactitude with which the initial identification is made.

Accordingly, we owe a double debt to Ullmann-Margalit for introducing a new category of explanation in her article, "Invisible-Hand Explanations,"¹ and for characterizing these explanations only in a rough-and-ready way which leaves many questions unresolved. My aim here will be to repay this debt in kind, if not in quantity: I will try to show that an attempt to define more clearly the nature of these explanations yields not one but three distinct forms of explanation. I should qualify this immediately, however, for Margalit distinguishes two varieties of "invisible-hand" (hereafter, I-H) explanations, and the refinements I introduce pertain to only one of them. So let me begin by saying what distinguishes I-H explanations from others, and what the two varieties of them are said to be.

* I owe thanks to Edna Ullmann-Margalit, Merrilee Salmon, Sandra Mitchell, and especially Carl G. Hempel and Joseph Camp, for their encouragement and useful comments on earlier drafts of this paper.
1 Synthese 39, 1978, pp. 263-291.

1: INVISIBLE-HAND EXPLANATIONS

Margalit provides a general characterization of I-H explanations in saying that they treat social phenomena (patterns and institutions) as "interposed between" the artificial and the natural realms, being the "result of human action but not of human design."² The phenomenon is accounted for as "a spontaneously formed order" arising in an unplanned way from the behavior of "numerous individuals, each busily doing his or her own private narrow bit."³ Thus, the coming to be or continuing to be of social phenomena are suitable I-H explananda. The character of I-H explanans is defined in an essentially negative way, as not involving appeal to a designer. We find a social arrangement which looks as if it might or must have (depending on our prejudices) come about through the guiding intervention of a designer who is, however, nowhere to be seen. The key to finding an explanation is to identify then a mechanism, the "invisible hand mechanism," which performs the work of social coordination that we might have been tempted to assign to the unseen hand of God or man. In Adam Smith, from whom Margalit has borrowed the idea of the "invisible hand," it is the mechanisms of the unfettered marketplace that coordinate the activities of individuals pursuing their private interests in such a way, it is said, as to secure the common good. The aim is to explain a pattern of productive and mutually beneficial interaction even in the face of perhaps no one's trying to benefit anyone other than himself.

There are, on Margalit's account, two distinct kinds of I-H mechanisms, so general characterizations of I-H explanations must

2 Ibid., p.263.

3 Ibid., p.271.

here come to an end. The first kind of I-H mechanism is one "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."⁴ This is the kind of mechanism that Smith had in mind, obviously, and which we shall explore after these preliminaries. Margalit conceives of the second form of I-H mechanism as a process of natural selection operating on social arrangements;⁵

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.

And so, for instance, the rain ceremonials of the Hopi may be "screened through" because they serve well the role 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."⁶

Invoking an evolutionary mechanism of this kind leads to a very different style of explanation from the "aggregate" I-H accounts which invoke an aggregating mechanism. The description of this pro-

⁴ Ibid., p.278. Stipulating that neither foresight nor intent can be present seems unnecessarily restrictive if the point is to rule out intentional design. Individuals in a freely competitive marketplace, for instance, might foresee the price of beets rising to a certain level, but be unable to influence the level of that price. Again someone might inefficaciously try to make it reach the level it in fact does. But such foresight and effort will simply be irrelevant to the (I-H) explanation we would give, since the relevant agents lack sufficient power to bring about what is to be explained. Apparently it will suffice to stipulate that no one individual or collective possesses all three of these factors (viz., power, effort, and foresight).

⁵ Ibid., p.282.

⁶ R.K. Merton, Social Theory And Social Structure, enlarged ed., NewYork 1968, pp. 118-119.

cess of natural selection presupposes that societies are systems in which institutions serve functional roles, and, following the standard literature,⁷ Margalit contends that the full-fledged (I-H) explanation is produced by conjoining the evolutionary account to an analysis of the explanandum's social function:⁸

an effort is made to find out [the explanandum's] contribution (if any) to the equilibrated 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.

Margalit calls this form of I-H explanation "functional-evolutionary" (hereafter, F-E), and suggests that there is room for both these and aggregate I-H explanations since they serve complementary, not conflicting, explanatory functions. Aggregate accounts, she says, provide "a chronicle of emergence,"⁹ whereas F-E accounts explain the "continued existence and prevalence"¹⁰ of the explanandum. This happy reconciliation -- indeed the very idea of F-E explanation in the social realm -- seems highly problematic to me for reasons that I will advance in section 4. Sections 2 and 3 I devote to an extended treatment of aggregate I-H explanations.

⁷ In her remarks on functional explanation Margalit takes herself correctly, I think, to be setting out the received view, insofar as there is one. Her references are to C. Boorse, "Wright on Functions," Philosophical Review 85, 1976, pp. 70-86; R. Cummins, "Functional Analysis," Journal of Philosophy 72, 1975, pp.741-765; W.C. Wimsatt, "Teleology and the Logical Structure of Function Statements," Studies in History And Philosophy of Science 3, 1972, pp. 1-80; and L. Wright, "Functions," Philosophical Review 82, 1973, pp. 139-168.

⁸ Ullmann-Margalit, op. cit., p.282.

⁹ Ibid., p.284.

¹⁰ Ibid., p.286.

2: THE GENERAL FORM OF AGGREGATE I-H EXPLANATIONS

Aggregate I-H explanations involve reference, as we have seen, to a mechanism "that aggregates the dispersed actions of individuals into the overall pattern..."¹¹ The full-blown account of this aggregating mechanism will begin, as Margalit explains it, with the description of an initial stage prior to the appearance of the explanandum, and will proceed through successive stages to a final stage where the explanandum is fully present. This initial stage will consist simply of individual persons with their individual intentions, beliefs, and goals (not to include any conception of the overall pattern ultimately produced), in a specified set of circumstances. Since the explanandum is to be explained as the result of the aggregated actions of individuals, and since these actions are to be explained in terms of beliefs and desires, it must be assumed that the participating individuals are rational. As Margalit correctly points out, this assumption of rationality has two components: the assumption of normalcy of beliefs and goals and the assumption that actions will be instrumentally rational with respect to those beliefs and goals.¹²

Very oddly, however, Margalit says the point of this rationality assumption is to guarantee that the story conveying the I-H process "sound like a description of the ordinary and normal course of events."¹³ Its sounding this way is made a condition for the story's constituting a well-formed I-H explanation. Margalit is driven to this, I think, because she believes that "it is the detailed stages of the invisible-hand process which ... supply

11 Ibid., p.287.

12 Ibid., p.288.

13 Ibid., p.271.

the aggregating mechanism."¹⁴ They supply it, she says later,¹⁵ in the sense that the process consists of those temporally ordered stages. Her predicament is that since she takes the specification of a mechanism to be nothing more than the serial listing of its stages, she needs a standard external to the mechanism itself which will guarantee continuity from one stage to the next. (It isn't enough to include at each stage the desires operating at that stage if no attempt is made to explain how the choices and conditions of one stage lead to the desires operating at the next.) The standard she chooses is this requirement that the progression of stages should give the impression of normalcy.

Surely it is the case, however, that the specification of any such mechanism should include an identification of the dynamic principles that explain the transitions from one "machine state" to the next. Recognizing this would yield an account that is much cleaner, while substantially identical with Margalit's. The occurrence of the individual actions that combine to form the explanandum may be explained as the result of forces -- directed desires, and surely these forces (together with conditions in the broader social environment which may themselves be changing) provide the impetus for, and determine the shape of, the transitions from one stage of the process to the next. The principles governing the operation of these forces are, I suggest, principles of rationality such as the following:

- (i) Normally, if A has good evidence that p, and it would be significantly useful to A to believe that p, A will believe that p.
- (ii) Normally, if q follows from p, and A believes that p, but p lacks some immediate import for action possessed

14 Ibid.

15 Ibid., p. 273.

- by q, A will also believe q.
- (iii) Normally, if A desires x and believes that y is his or her best means to securing x, all things considered, then A will (lacking y) also desire y.
- (iv) Normally, if A desires x and believes that his or her best (all things considered) manner of securing x is the performance of action, a, A will (all else being equal) do a.

As we've already seen, Margalit must make use of assumptions about the rationality of beliefs, goals, and actions, and so the departure from her account will not be too great if these are mobilized for explicit use.

In Margalit's most fully outlined example, Schelling's model of segregation,¹⁶ it is quite clear that the motives of the individual participants explain not only the transitions from one step to the next but also the stability of the final stage. De facto segregation is explained, in somewhat idealized fashion, as the result of individuals desiring to live in locations where "their own color group is not in a minority in their immediate neighborhood,"¹⁷ it being assumed that those who are dissatisfied will pursue this desire in an economical way, namely by moving to the nearest location where their desire will be satisfied. A neighborhood at large is represented by an axis whose points represent houses, and the "immediate neighborhood" of a house is defined as the house itself together with the first four houses in each direction from it. The initial stage consists of 70 individuals, 35 white and 35 black, distributed one to a house and randomly with respect to color. It turns out that only two rounds of moving are required to reach a situation in which everyone is satisfied, the unintended result being six stable segregated clusters. Both of

16 T.C. Schelling, "Models of Segregation," American Economic Review 59, 1969, pp.488-493.

17 Margalit, op. cit., p.272.

these transitions are adequately explained by the stipulated desires, together with the situations in which the bearers of those desires find themselves. What the details of each stage depict is the progress towards the final result; what they, together with the principles of rationality, explain is the choices made by participants in pursuit of their desires. These choices determine the circumstances of the next stage, and those circumstances in turn explain, together with the participants' enduring desires and the principles of rationality, the specific (proximate) desires and choices that emerge at the next stage. The emergence of many social patterns might differ from this, I take it, only in involving more interesting interactions among the participants, and in the circumstances at various stages being dependent on changes in addition to those resulting from the choices made at prior ones.

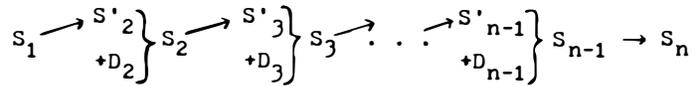
Margalit says that "the full-fledged description of the invisible-hand process falls under Hempel's category of genetic explanations,"¹⁸ though she does not pursue this claim far enough to see what follows from it. Hempel outlines the form of genetic explanations as follows:¹⁹

... schematically speaking, a genetic explanation will begin with a pure description of an initial stage; thence, it will proceed to an account of a second stage, while the balance is simply added descriptively because of its relevance for the explanation of some parts of the third stage, and so forth.

The following diagram schematically represents the way nomological explanation is combined with straightforward description in a genetic account of this kind:

18 Ibid., p.270.

19 C.G. Hempel, "Aspects of Scientific Explanation," in his own Aspects of Scientific Explanation, New York 1965, pp.449-450.



The arrows indicate nomic connections between stages, and "S₁, S₂, ..., S_n are sets of sentences expressing all the information that the genetic account gives about the first, second, ..., nth stage," each except S₁ and S_n being composed of some sentences, S'₂, S'₃, ..., S'_{n-1}, which specify features "explained by reference to the preceding stage," and some D₂, D₃, ..., D_{n-1}, which give additional unexplained facts. This is the general form that genetic explanations take, but there are special cases, such as Schelling's model, in which no additional unexplained facts need be called upon. The schema then is simply S₁ → S₂ → S₃.

Margalit fails to follow up this suggestion that aggregate I-H explanations are genetic explanations, because, it appears, she is reluctant to identify the form of the former in a precise way.²⁰ Lacking coherent grounds for such reluctance, there are important matters to be settled, paramount among them the issue of what will provide the nomological interconnections between different stages. One suggestion, though certainly not an unproblematic one, is to construe the principles of rationality stated above as empirical statistical laws of human nature. This would open the way for explaining the individual actions which effect transitions from prior stages to latter, though these laws would not be the only ones required in moving from S₁ to S'₂, S₂ to S'₃, and so on. This is so because we would be making inferences not only from

20 "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." Ullmann-Margalit, op. cit., p.270.

situations to beliefs, and from beliefs and desires to courses of action, but also from initial situations and courses of action to consequent situations. This last kind of move presents complications because appeals to agent's motives may account for what they set out to do, but what they set out to do is not always what, nor ever all that, they succeed in doing. Sentences in S' sets will capture, among other things, what agents have succeeded in bringing about.

3: TAKING HOW'S AND WHY'S SERIOUSLY

As a preliminary to any further progress in clarifying the form of these aggregate I-H explanations, we must now consider the central explanatory functions they are claimed to serve. In addition to providing accounts of the genesis of social patterns, Margalit maintains that they explain how and why the patterns are perpetuated. They "contribute to our understanding of the inherently self-reinforcing nature of [given patterns] and hence of [their] being successful and lasting," she says.²¹

How, we must ask, can an aggregate I-H explanation, which so far has been characterized as a kind of genetic explanation, explain how and why a social pattern continues to exist? The final stage of one of these explanations will present in some form a great many facts about the components of the explained pattern, but given

21 Ibid., p.275. Margalit claims that this is true of all well-formed aggregate I-H explanations, true and false alike. She also claims that even false ones provide rational reconstructions and reductions of the concepts of their explananda. Though interesting, these contentions are all easily shown to be mistaken, I think, and considerations of space incline me to press on with the main line of inquiry.

what has so far been said, these facts need not, and often will not, take the form of an account of the structure ("the inherently self-reinforcing nature") of the explanandum. How are the descriptive statements at S_n organized? Surely we cannot allow just any compilation of descriptive sentences to count as identifying the nature of a pattern. In Schelling's model we can just see how the individual components add up to the explained pattern, but that's because the information is represented in a particular way that is not dictated in the least by its being an aggregate I-H explanation. The information is recorded one entry at a time, but in such a way as to automatically add up to a picture of the explanandum. Most one-entry-at-a-time recordings of the activities of individuals will not automatically yield a profile of any pattern formed by those activities, but will require interpretation. Tracing the genesis of a social arrangement will often facilitate such an interpretation, but the genetic account will not itself be such an interpretation.

We need, I suggest, an entirely distinct form of explanation to do the job of explaining how the structure of a pattern or institution makes it "inherently self-reinforcing." Aiming for a form of explanation which departs as little as possible from the materials already at hand, my suggestion is that we conceive of the explananda as systems whose behavior is to be accounted for in terms of the interaction of a number of functional components. Let's call this systematic explanation,²² and take as a model the explanation of

²² after J. Haugeland, "The Nature And Plausibility of Cognitivism," The Behavioral And Brain Sciences 2, 1978, p.216. Archaeologists, among others, would probably call this "the systems approach" to explanation and contrast it in their work with the older "diffusionist" model of cultural development.

the behavior of an automobile engine in terms of the coordinated contributions of its components: the carburetor, ignition system, and so on. These components can be specified simply as input-output functions ("black-boxes") and arranged either discursively or diagrammatically so as to indicate how the behavior of the system as a whole results from their functional integration. This "arranging" amounts to the proposition that the products of each component are handed over to the components that take those as their input. That suffices to explain how the system works, how it can produce what it does (locomotion), given what is available to it (air, gasoline, smooth roads, etc.). Some of the system's components may themselves be systems (thus, subsystems of the engine), in which case their behavior, originally captured simply as law-like functional (in the mathematical sense) dependencies of output on input, may in turn be systematically explained at a second stage of analysis.

The strategy is perhaps not so different from how we explain the validity of a multi-lined formal proof. At each step we appeal to a rule of inference, a kind of input-output function, to justify setting down a new formula (output), given specified prior lines of the proof (input), and we proceed from start to finish so as to certify that (and see how) it carries us along from the original premisses (input to the system as a whole) to the final conclusion (output of the system as a whole). Doing this brings us to an understanding of how the "inference machine," composed of those inference rules, can yield the product it does given the materials at its disposal. It analyzes a superfunction, so to speak, into a structure of component subfunctions. Since specifying the functional

components of the "inference machine" is just identifying the rules of inference, and since there is nothing more to proving validity than showing that every line is indeed justified under a truth-preserving rule, then, to make explicit the implications of the analogy, there isn't anything more to giving a systematic explanation than specifying the functional components and getting the connections between them right, where getting the connections right is being sure that inputs, unless they are inputs to the system as a whole, are shown as coming from the components of which they are outputs. Similarly, since the activity of each component is captured as a law-like function, there is a strong nomological element to these explanations.²³

To clarify now why these systematic explanations might provide what Margalit needs, let us recall that the point of using the title "invisible-hand" is to contrast these explanations with ones from intentional design in just those cases where there is a temptation to think explanations from intentional design appropriate. Such a temptation exists exactly in those cases where the social pattern or institution has a structure so complex or highly developed that it would be surprising if the coordination of individual agents could have come about without planning. Put more directly, the title "invisible-hand" is appropriate only where the explanandum is a relatively complex social structure, and where there is at least "a difference in type between the overall pattern to be

23 I do not, in using this analogy, mean to imply that the components of all systems are serially arranged. Indeed, the attentive reader will note that the lines of the proof are serially arranged, but the "inference machine's" components are not. (Cf. the lines of a computer program vs. the organization of its subroutines.)

explained and the individual actions which are supposed to bring it about,"²⁴ if not several interacting types of actions different in type from the pattern.

Consequently, most of the social phenomena for which I-H explanations are appropriate should be complex in the sense that they arise out of the interactions of a large number of individuals whose behavior can be grouped into several disjoint similarity classes. When such classes can be identified the chances are good that the pattern can be explained systematically. In Margalit's example of the creation of money in the banking system²⁵ we can identify three basic groups: those who save, those who borrow, and those, the institutions, which mediate between savers and borrowers. Spelling out the behavior of these groups as they interact with one another serves to explain how the system does what it does, and thereby gives us insight into the nature of the phenomenon. In Schelling's model of segregation everyone is doing the same thing; no distinct groups can be identified, and so no systematic explanation is available. This is simply a symptom, however, of the fact that, plausible as the genetic account of it may be, de facto segre-

24 Ullmann-Margalit, op. cit., p.261.

25 Ibid., p.264: "No one needed to have invented the commercial banking system, nor need anyone have invented 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."

gation (on this model) is not complex in a way that makes it a representative object of I-H explanation.

Systematic explanations are not all we will require, however, if we're to understand what keeps arrangements going -- if we're to see them as "inherently self-reinforcing." If the question is "Why is the pattern perpetuated?" we need first and foremost to understand why people in sufficient numbers will do what members of the component-classes do. Our explanation must identify incentives that attach to playing the roles that constitute component-class membership, and show how these incentives suffice to maintain a level of participation sufficient for the perpetuation of the arrangement. This will involve dispositional claims about human nature as well as facts about the number of potential component-class members and the conditions in which they are making their decisions. The strategy is to show that a certain percentage of people who could potentially fill the roles in question will, given the range of choices, choose to participate, and furthermore that, given the number of such potential participants, that percentage will be enough. It will also usually involve showing how it is possible for the structure of incentives to accommodate various changes in environmental constraints. The notions of regulatory and stabilizing mechanisms may be invoked here, but such mechanisms, when they are implicated, will already have been counted as functional components under the systematic explanation of the phenomenon.

In the example at hand, the explanation begins by identifying the reasons people have for saving, for borrowing, and for running and working in banks. The operation of banks could itself be given a systematic explanation, and this, together with a profile of the

environmental constraints on their operation, will explain how it is possible for the system of incentives to remain intact and effective. Many social phenomena are not so well-defined, of course, but even for them there may often be explanatory utility in developing idealized systematic models, a maneuver which is by no means without precedent. Where this strategy fails it is not at all clear that Margalit's claims for the explanatory power of aggregate I-H explanations can be redeemed.

In trying to clean up Margalit's account of aggregate I-H explanations we have now distinguished a total of three forms of explanation: the genetic, the systematic, and the dispositional explanations which build upon the systematic. The first two of these, or better, the first and a hybrid of the second and third, I count as distinct forms of I-H explanation. Programmatic as these suggestions are, they represent a significant refinement of Margalit's account.

But here it may be objected that I have given the why question short shrift in suggesting that it may be answered by appeal to the motives and circumstances of those who participate in the practice to be explained. For sometimes, it will be pointed out, the question that really interests us looks beyond the circle of participants to the possibility of outside intervention. "Why," we ask, "do the American people allow such a state of affairs to continue?" ("Why the lack of intervention?") Or again, "Why did Congress create new incentives to stabilize participation in this threatened institution?" ("Why the intervention?") Answers to questions of both these varieties may cite the influence of private interests on those who might intervene. In other instances they may cite false or true

beliefs about (and a concern or lack of concern to advance) the common good. Failures to intervene may also be explained by reference to a lack of awareness of the situation, or lack of the power, resources, or organization essential to making a difference.

This seems to me not so much an objection, however, as simply an interesting complication. Explanations that make reference to outside intervention or its absence may supplement, but never supplant, ones based on the motives of, and patterns of interaction among, participants, for understanding the internal dynamics of a pattern or institution is plainly more central to understanding that pattern or institution. Furthermore, the explanatory relevance of intervention will in many instances provide us with a broader I-H explanation than we would have constructed otherwise. In some of these cases the additions will be integral to a larger systematic scheme, whereas in others their import from the systematic point of view will merely be that they don't interfere with the rest. In this latter sort of case, after identifying the relevant classes of actors and charting their interactions, we would add a further class (viz., those who are not interfering, but might be expected to) whose members go about their own private affairs in preference to interfering, or lacking information that would incline them to interfere. The former sort of case may arise when the behavior of those who might or might not interfere must be bought through payoffs, threats, or deception.

This may be illustrated by the pattern (striking frequency) of handgun murders in the United States today, a phenomenon which lends itself admirably to systematic analysis. We begin with a puzzle: given that manufacturers and retailers can only (legally)

put registered guns in the hands of potential killers, and given also that those who intend to use guns in the commission of crimes strongly prefer ones that are not registered in their own names, how is it that so many murders involve guns that cannot be traced (through registration) to their users? This puzzle is then resolved by identifying a mechanism which takes traceable guns as input, and transforms them into untraceable ones. Two classes of actors are implicated here: those moved by a concern for their own safety to purchase firearms and conceal them ineffectively, and those who steal them, usually in burglarizing houses.²⁶ These links close the self-reinforcing cycle in a way which largely accounts for the level of violence to be explained: violence generates fear, and that fear, in turn, generates a steady supply of untraceable handguns which may be put to illicit use. Moreover, the danger to aggressors is also heightened in such a way as to encourage them to use greater force. This explanation may come to seem incomplete, however, when we consider how easily this cycle might be broken by appropriate legislation. Our investigation of Congressional and popular inaction might then lead to a broader systematic picture of how political and economic power are maintained, depending on how much significance we attached in the end to industry influence through public relations efforts, lobbying, contributing to political campaigns, and so on.

A final important observation I must make regarding these supplemental explanations is that in explaining the existence of

²⁶ While I have not seen the relevant studies, police investigators for the city of New Orleans inform me that in nine of every ten instances in which handguns are involved in crimes, they are used by persons other than their lawful owners.

social phenomena, only a narrow subclass of them will give a decisive role to the fact that the phenomena contribute to social well-being (i.e., have social functions). Among these some will be I-H (by my lights, perhaps not by Margalit's; see footnote 4) because the intervening or non-intervening individuals act from an informed desire to promote the common good, but lack the power or organization to count as social designers. (Again, there is no difference of great significance between this and the (rare) case in which some or all of the participants in a pattern act in ways motivated by the common good, but independently of one another, and with no control over what the others do.) In cases where true beliefs about social function are mediated by the agency of some individual or collective having designer status, the explanation will not be I-H, of course. Rather, the explanation will be much like that of the existence of an artifact which serves the purpose it was intended to.

4: CONCLUDING REMARKS ON THE FORMS AND FUNCTIONS OF I-H EXPLANATIONS

Let us now recall that Margalit takes there to be two forms of I-H explanation, aggregate and F-E, and that she claims they address different questions and so can serve complementary explanatory roles. To repeat, aggregate accounts are supposed to provide "a chronicle of emergence," whereas F-E accounts are supposed to explain the "continued existence and prevalence" of explananda by citing their raisons d'être. We have just seen, however, that (as Margalit has maintained throughout) aggregate explanations can explain how and why social phenomena persist. (Her apparent inconsistency on this point is baffling.) Consequently, these explanatory functions cannot

be distributed in quite the tidy manner Margalit envisions.

To determine how they should be distributed we must consider whether F-E explanations can do what Margalit says they can. I will argue, first, that they could not be so broadly available nor easily generated as Margalit suggests, since they depend on three assumptions that are not generally reliable. A closer inspection of the third of these assumptions will then show that the idea of an F-E, I-H explanation is untenable. The effect of this is to narrow the field of I-H explanations to those elaborated in sections 2 and 3: the genetic and the systematic-dispositional.

Functional-evolutionary explanations depend, first, on the pattern or institution's having a function in the sense of contributing to social well-being. No doubt there are many institutions that make social contributions, but there are probably far more patterns and institutions that do not, as my example of the pattern of handgun murders illustrates. Margalit does explicitly admit the possibility of an institution's having no social function, but she uncritically admits the biological model into the realm of social explanation. We should notice, first, that even in the biological realm we cannot assume that selective pressure is so intense as to insure that every anatomical structure has a life-sustaining function. Secondly, there are disanalogies between the biological and social realms that should make us even less sanguine about finding functions for social explananda. One such disanalogy is the fact that persons are more autonomous than tissue; they are able to act in ways that suit their own personal requirements irrespective of what society might require. Consequently, in trying to account for a social arrangement we are on much safer ground in assuming that

the individuals who participate in it act for reasons, than we are in assuming that the arrangement contributes to the "equilibrical and frictionless survival" of the host society. Aggregate explanations will, therefore, be more widely applicable.

Again, even when we can attribute a function to a social arrangement we cannot assume, as Margalit does, that even its "faint beginnings" served the same function that it does now. Margalit talks as if the whole evolutionary development of the explanandum will just fall out of the analysis of current function, but that cannot be the case. Some institutions are able to survive precisely because they manage to adapt themselves, in a changing environment, to entirely different functions. In such cases the explanation would be incomplete without histories of the changes in both function and the environment to which the institution adapted. So to identify the item's current function can scarcely be to have "all but explained" it (its presence).

Finally, even when we can attribute a function and assume sameness of function over time, we may be wrong in assuming that the explanandum has evolved in any sense that warrants claiming that an "evolutionary mechanism" has provided the functional analysis with causal teeth. My concern now is not, as earlier, with the assumption of selective pressure, but rather with what might be called the assumption of diversity. There must, as Margalit says, be an "inventory" of social arrangements from which ones well suited to given roles can be selected. If there is no range of candidates to select from, then there can be no process of selection, and consequently no explanation from natural selection -- i.e., no explanation that is evolutionary in the relevant sense. Again, perhaps some social arrangements have been selected for

their roles from fields of candidates, but are we really to believe, for instance, that Hopi rain dances triumphed over and against competing, diverse practices which disappeared because they were less effective in reinforcing group identity? Or did it win out over other practices only in the sense that those who came to practice it found themselves with less time and energy to devote to other things? Again, taking examples which may be explained by reference to individual dispositions, though additively (like segregation on Schelling's model) rather than systematically, what about widespread institutions like marriage and living in single-family dwellings? As before, a serious disanalogy with the biological realm emerges: in the social realm there are two important forms of preselection which reduce the significance of actual selection. On the one hand, people, like other social animals, have heritable dispositions to act in certain (often socially enhancing) ways rather than in others. To some extent, then, there is an indirect biological preselection which restricts the range of social arrangements which may arise. On the other hand, unlike other social animals, people are intelligent enough to sometimes figure out that certain arrangements won't work, without having tried and suffered the consequences. These considerations suggest that the assumption of diversity can only be made selectively, and on the basis of historical investigation.

The general lesson that emerges here is that true F-E explanations must be both rarer than Margalit implies, and more laboriously acquired. Beyond this, however, there are conclusive reasons for denying F-E explanations the status of a distinct class of I-H explanation. To see this we must ask what explanatory force the

appeal to an evolutionary mechanism has in those cases where the relevant assumptions are secure, and the explanandum's serving some function has made the difference in its continuing to exist. In saying that its serving a function has made this difference we view the explanandum, of course, against the comparison class of other diverse phenomena which did not serve that function efficiently and were eliminated by environmental pressures. In an important sense, however, what has happened to these other phenomena is quite irrelevant to the explanandum's surviving; from a more local perspective it has, through whatever means, simply been "lucky" enough to be so constituted as to survive those external pressures which the others didn't survive. Again, considering the kinds of causal relevance that persons may bear to the explanandum, we can say that this survival may be either fortuitous or, on the other hand, the result of successful social design (i.e., not fortuitous). That is, the arrangement's having a certain structure and serving some social function may have resulted in a way that no one controlled (or could have controlled) from various individuals going about their private affairs (with or without any relevant concern for the common good). In this case the causal story to be filled in is an aggregate I-H account. If, on the other hand, it is the planned result of intentional intervention, then the explanation is no kind of I-H explanation at all.

Thus, the evolutionary element in F-E explanation seems merely to stand in for two distinct and familiar kinds of mechanisms, and so to do no explanatory work of its own. The F-E model is untenable in the social realm, therefore, and so Margalit's attempt to distribute the explanatory functions of I-H explanations between

aggregate and F-E accounts collapses completely. Moreover, these two modes of causal mediation by which function can enter are just what we found them to be at the close of section three, except that now, when the evidence warrants it, we may add the non-intervention of a potentially lethal environment as a supplemental account on a par with other explanations by non-intervention. But this, I hasten to point out, will be a kind of causal account which in its own right can make no claim to identifying a social phenomenon's reason for existing. (Providing raisons d'être was said to be the explanatory function of F-E accounts, recall.) Since social arrangements cannot properly be said to have reasons at all, the reasons that sustain them can only be those belonging to the individuals who participate in or intervene (refrain from intervening) on their behalf.

One final remark is in order. If I have tried to minimize the significance of a social phenomenon's function in the context of explaining its own existence, I have also tried to show in setting out my systematic-dispositional model that functional analyses of social items figure importantly in explaining the workings and capacities of systems in which those items are components. Margalit closes her paper emphasizing the distinction between aggregate and F-E I-H explanations, and so it is ironically that I now close responding not only that F-E accounts are no kind of I-H explanation at all, but also that the primary place for functional analyses in the realm of social explanation must be in the guise of a kind of aggregate account.