

SYNTHESE

An International Journal for Epistemology, Methodology and
Philosophy of Science

Editors:

JAMES H. FETZER
Dept. of Philosophy, Univ. of Minnesota
Duluth
MN 55812, U.S.A.

RISTO HILPINEN
Dept. of Philosophy, Univ. of Turku
SF-20500 Turku, Finland

BARRY RICHARDS
Imperial College of Science, Technology and
Medicine, 180 Queen's Gate,
London SW7 2BZ, U.K.

Editorial Board:

Robert F. Butts, *Univ. of Western Ontario, London, Ont., Canada*; Nancy Cartwright, *Stanford Univ., Calif., U.S.A.*;
John Echeburri, *Stanford Univ., Calif., U.S.A.*; Anne Fagot-Lalonde, *Paris, France*; Jens Erik Fensholt, *Univ. of Oslo,
Norway*; Kit Fine, *Univ. of California, Los Angeles, Calif., U.S.A.*; Dagfinn Føllesdal, *Univ. of Oslo, Norway*; Marjorie
Greene, *Univ. of California, Davis, Calif., U.S.A.*; Robert Howell, *State Univ. of New York at Albany, N.Y., U.S.A.*; Hide
Ishiguro, *Kyoto University, Tokyo, Japan*; Ernest LePore, *Rutgers Univ., New Jersey, U.S.A.*; Barry Loewer, *Rutgers
Univ., New Jersey, U.S.A.*; C. David McCarthy, *C.U. Moulines, Freie Univ. Berlin, Germany*; Brent Mumby, *Syracuse
Univ., New York, U.S.A.*; Ilkka Niiniluoto, *Univ. of Helsinki, Finland*; David Pearce, *Freie Univ. Berlin, Germany*; David
Pearce, *Christ Church, Oxford, U.K.*; Vadim Sadovskiy, *Academy of Science, Moscow, Russia*; Matti Suuronen, *Univ. of
Helsinki, Finland*; David W. Smith, *Univ. of Calif., Irvine, U.S.A.*; Joseph D. Steele, *Colorado School of Mines, Golden,
Colo., U.S.A.*; Goran Sundholm, *Leyden University, The Netherlands*; Patrick Suppes, *Stanford Univ., Calif., U.S.A.*;
Richardson H. Thomason, *Univ. of Pittsburgh, Pa., U.S.A.*; Jan Wolenski, *Jagiellonian Univ., Cracow, Poland*.

Honorary Consulting Editors:

Arthur W. Burks, *Univ. of Michigan, Ann Arbor, U.S.A.*; Jerry Fodor, *CUNY, New York, U.S.A.*; Ian Hacking, *Univ.
of Toronto, Canada*; Esa Saarinen, *Univ. of Helsinki, Finland*; Merrill H. Salmon, *Univ. of Pittsburgh, Pa., U.S.A.*;
Wesley C. Salmon, *Univ. of Pittsburgh, Pa., U.S.A.*; Anos Tversky, *Stanford Univ., Calif., U.S.A.*; Thomas Wasson,
Stanford Univ., Calif., U.S.A.

Publication programme, 1993: Volumes 94-97 (12 issues).

Subscription prices, per volume (3 issues): Dfl. 306,—/\$ 191.50 including postage (Dfl. 1224,—/\$ 766.00 per annum).
Subscriptions should be sent to **Kluwer Academic Publishers Group, P.O. Box 322, 3300 AH Dordrecht, The Netherlands,**
or at **P.O. Box 358, Accord Station, Hingham, MA 02018-0358, U.S.A.**, or to any subscription agent. Private subscriptions
should be sent direct to the publishers.

Changes of mailing address should be notified together with our latest label.
For advertisement rates, prices of back volumes, and other information, please apply to Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

*Kluwer Academic Publishers incorporates the publishing programmes of D. Reidel, Martinus Nijhoff, Dr W. Junk and
MTP Press.*

Photocopying. In the U.S.A.: This journal is registered at the Copyright Clearance Center, Inc., 27 Congress Street,
Salem, MA 01970.

Authorisation to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted
by Kluwer Academic Publishers for users registered with the Copyright Clearance Center (CCC) Transactional Reporting
Service, provided that the fee of \$5.00 per copy is paid directly to CCC. For those organisations that have been granted
a photocopy licence by CCC, a separate system of payment has been arranged. The fee code for users of the Transactional
Reporting Service is 0039-7857/93\$5.00.

Authorisation does not extend to other kinds of copying, such as that for general distribution, for advertising or promo-
tional purposes, for creating new collective works, or for resale.

In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to Kluwer
Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Microfilm and microfiche editions of this journal are available from University Microfilms International, 300 North Zeeb
Road, Ann Arbor, MI 48106, U.S.A.

1993 Kluwer Academic Publishers. Printed in the Netherlands.

SYNTHESE / Volume 94 No. 2 February 1993

WILLIAM S. BOARDMAN / The Relativity of Perceptual Knowledge 145

ELLIOT SOBER / Temporally Oriented Laws 171

GEOFFREY BRENNAN and PHILIP PETTIT / Hands Invisible and
Intangible 191

DANIEL M. HAUSMAN / Why Don't Effects Explain Their Causes? 227

P. P. ALLPORT / Are the Laws of Physics 'Economical with the
Truth'? 245

NANCY CARTWRIGHT / Is Natural Science 'Natural' Enough? A
Reply to Phillip Allport 291

PIERS RAWLING / Choice and Conditional Expected Utility 303

Announcement 329

HANDS INVISIBLE AND INTANGIBLE

ABSTRACT. The notion of a spontaneous social order, an order in human affairs which operates without the intervention of any directly ordering mind, has a natural fascination for social and political theorists. This paper provides a taxonomy under which there are two broadly contrasting sorts of spontaneous social order. One is the familiar invisible hand; the other is an arrangement that we describe as the intangible hand. The paper is designed to serve two main purposes. First, to provide a pure account of the invisible hand, with some indication of the varieties of invisible hand (and, indeed, backhand) available. Second, to develop and motivate the unfamiliar conception of the intangible backhand. We believe that a recognition of the availability of this latter sort of spontaneous organising mechanism – and the mechanism is implicitly recognised in many traditions – is of great importance in political theory; it is of particular importance nowadays when the usual focus is entirely on the invisible hand.

INTRODUCTION

Our aim in this paper is to distinguish and characterise two devices whereby social order may be spontaneously achieved. One is the familiar invisible hand; the other is something that we describe as the intangible hand.

The notion of a spontaneous social order, an order in human affairs which operates without the intervention of any directly ordering mind, has a natural fascination for social and political theorists. It suggests the possibility of parsimonious explanation, with institutions that bear the marks of intentional control yielding to accounts that invoke no controllers. And it holds out the prospect of ways of regulating social life which are proof against the fallibility and corruptibility of regulators.

But the notion of a spontaneous order is ambiguous and we should immediately make it clear what exactly our topic is. The spontaneous order may be taken as an order that arises spontaneously or an order that maintains itself spontaneously. Our primary interest is in the spontaneous order in the second sense: the order that is self-sustaining, whether or not self-starting; the order, as it is often put, which is self-regulating. Some such spontaneous orders may have arisen spon-

taniously; others will have arisen with the help of legislators and administrators.¹

One further ambiguity also calls for clarification. The notion of an order may refer us to any robust pattern in human affairs or just to one that is suitably valued: disvalued orders would be forms of disorder on this narrower usage, where they would remain forms of order on the broader one. We are interested in both valued and disvalued arrangements – arrangements produced, for example, both by invisible hands and invisible backhands – and it seems appropriate to regard them both as forms of order. Thus we take the notion of an order to be the broader one that refers to any robust pattern in human affairs.

There are two propositions we wish to defend in the paper. The first is that the spontaneous ordering mechanism usually described as the invisible hand (or backhand) can be usefully characterised by the behavioural controls that it deploys: it is any sort of mechanism which mainly supports the order in question by means of controls that are decentralised and, as we shall say, contingently non-intentional. The second proposition we defend is that there is another important spontaneous ordering mechanism – the intangible hand (or backhand) – and that this is characterised by its use of controls that are decentralised, as in the invisible case, but now essentially rather than contingently non-intentional. We think that the first proposition will generally be found of taxonomic interest, since there is surprisingly little written in abstract characterisation of the invisible hand (but see Nozick 1974; Ullmann-Margalit 1978). But it is the second proposition that we expect to command more attention. We believe that the intangible hand is a device of the greatest interest and we hope we can convey our sense of its importance.

The paper has three main sections. In Section 1 we present the concepts and distinctions that are later used in discussing the two sorts of social hands. In Section 2 we present our claim about invisible hand mechanisms. And in Section 3 we present our thesis about intangible hands. The paper concludes with a short comment on the history of thinking about these devices.

1. BACKGROUND

Any ordering mechanism in society depends on the operation of some distinguishable behavioural controls or influences: factors, that is,

whose presence makes it more likely that agents will choose that set of options which produce the pattern of social outcomes involved in the order. As we see it, there are two kinds of behavioural controls relevant in principle to our purposes – sanctions and screens – though we shall mainly be discussing sanctions. We do not claim that there are no other forms of behavioural influence, that sanctions and screens exhaust the set of possible behavioural controls. All we say is that they are significant forms, and that they are the relevant ones for the kinds of ordering mechanisms with which this paper deals.

What makes sanctions and screens relevant in particular among the full range of behavioural determinants is that they are controls which people exercise over one another. Not only do they serve as motors, like all controls, in the production of human behaviour; they serve as levers which can be manipulated by other human beings (Brennan and Pettit 1991). We can hope to design or redesign institutions of human interaction with a view to affecting the pattern of sanctions and screens which influence agents. We can hope to shape the institutions with a view, ultimately, to securing desirable rather than undesirable aggregate outcomes.

Sanctions operate on the set of options before an agent, making some options more attractive or less attractive than they would have been had the sanctions not been in place. Screens, by contrast, operate on the set of actors: they filter members of the group in such a way that those whose behaviour remains relevant to determining the particular social outcome are more likely to choose in the way required for that social outcome to emerge. Screens may operate independently as when they filter out of the relevant group those actors who do not have a natural predilection to choose the required action; or screens may operate in conjunction with sanctions by fixing the constituency in favour of those agents who are likely to be particularly susceptible to the sanctions applied.

Sanctions come in both negative and positive forms, either as penalties or rewards. The negative sanction will penalise the failure to choose appropriately; the positive sanction will reward the appropriate choice. Sanctions are clearly in operation when we can identify an intentional, sanctioning agency. In such a case we focus on the situation that would obtain in the absence of the agency and we see the responses of the agency to the doings of the relevant parties as attempts to affect those doings by imposing suitable rewards and penalties. But sanctions need

not involve the presence of any such sanctioning authority. It does not matter if there is no single agent or agency which imposes the rewards and penalties. And it does not matter if the rewards and penalties are not imposed intentionally. The only requirement is that people respond to the agent in such a way that it becomes more likely that he will choose one sort of option. They offer a reward for the appropriate choice and/or a penalty for the inappropriate one.

Sanctions in this broad sense will usually be recognised by the parties whom they affect, and may be internalised in their deliberations: the parties may reason that the option supported by the sanctions is the one to choose, given the reward it brings with it or the penalty it avoids. But sanctions may also operate without featuring as reasons in this way and without even entering into the consciousness of the agents they affect. Suppose that as things stand in a certain situation agents already have reason enough to choose a particular option but that sanctions are introduced, by whatever instrumentality, to support that option. Even if the agents in question do not become aware of the sanctions, still the sanctions may operate to make the choice of that option more likely than it would otherwise have been. They may mean that were the agents' actual reasons to shift away from the option, so that agents began to choose otherwise, then the agents would generally become aware of the reward forgone or the penalty suffered and would be returned to their original choice. Thus sanctions may operate to make the choice of a certain option more likely without figuring explicitly in the agent's deliberations. Sanctions may serve as standby supports that do not actually have an impact on people's decisions but that would do so were the decisions to go the wrong way; they may serve as counterfactual rather than actual controls. In such a case they will not explain the occurrence of those decisions, but they will explain why it is more or less inevitable that the decisions are of the relevant type: they program for the resilience of decisions of that type, ensuring that even if existing motivations failed, still such decisions would be forthcoming (Jackson and Pettit 1992a, 1992b; Pettit, forthcoming, Chs. 5 and 6).

There are many examples of sanctions which are not productive in the sense of actually figuring in deliberation. For most of us the sanctions of the criminal law do not figure in keeping us off the paths of crime; we deliberate in terms that make crime, or at least certain sorts of crime, more or less unthinkable (Braithwaite and Pettit 1990). But that does

not mean that these sanctions are irrelevant to action. On the contrary, the sanctions may still serve to make it more likely than otherwise that we will abide by the law. It may even be that were the sanctions not there and were we to become aware that we could commit certain crimes with impunity, then the reasons which lead us at present to shun crime would be trumped by considerations of criminal profit. Consider Glaucon's challenge in Plato's *Republic*. How far would someone who is just and law-abiding at present remain capable of resisting crime if he could use the ring of Gyges to make himself invisible at will?

The notion of the sanction is familiar to all of us in social life; we need only think of the criminal law to see obvious examples. The notion of the screen may be less familiar; but it is equally simple. Suppose, as is sometimes said, that academics are status-oriented. One possible explanation invokes a screening mechanism. In academe there is great scope for comparison between individuals, even individuals across different countries, and so the profession attracts those who care for being ranked by and among their peers, those who prize status. Others are screened out, as we say, or at least the likelihood of their entering the profession is reduced by this screening mechanism. The screening involved in this example is of a self-administered variety. But of course screens may also be applied by some individuals to others. The mechanism is used in the selection of employees, for example, as employers seek out the most promising candidates.

We must now introduce the taxonomy of behavioural controls which we believe necessary for the characterisation of invisible hand mechanisms and for the demarcation between invisible and intangible hands. We shall generally limit ourselves to examples of sanctions in developing this taxonomy, and later in applying it. We do this for simplicity's sake. We reckon that, with a little imagination, the taxonomy can be extended to cover screens as well.

Our taxonomy is built up out of three distinctions. The first distinction is one between centralised and decentralised sanctions. The centralised sanction is the sort of penalty or reward that requires action on the part of some central sanctioning body, such as an arm of the state. The decentralised sanction is the penalty or reward that comes to a person without any such central initiative: say, in virtue of the spontaneous responses of her fellows to how she is perceived to act. There may well be a grey area between the two sorts of sanctions, but it will be possible to place most examples clearly on one side or other of the divide.

Centralised sanctions will include, in the positive camp, the remuneration or promotion of public servants and the conferring of public honours on selected citizens; in the negative camp, they will include the rigours of the criminal law, such as the fines and terms in prison that the courts may impose. Decentralised sanctions will include the positive rewards of financial gain, honour, and influence and the negative penalties of financial loss, disgrace, and neglect.

The second distinction involved in our taxonomy is one between intentional and non-intentional sanctions. This hardly applies in the centralised case, as the responses of a central agency will almost always be intentional. But it does certainly apply with decentralised sanctions. To sanction someone intentionally will be to recognise the contribution their behaviour makes to the relevant result and to reward or penalise the agent for making that contribution. It is clear that there is often decentralised, intentional sanctioning. People may praise or admonish one another, for example, in planned and conscious approval or disapproval of a certain form of behaviour. But it is equally clear that people may impose sanctions, even impose them systematically, without the will or consciousness required for intentional sanctioning. Think of the way in which someone who tries to sell certain goods above the market price is penalised by consumers. They will penalise her by taking their custom elsewhere, but they may penalise her non-intentionally, lacking any sense that her behaviour is inimical to competitive pricing. And, even if they are aware that it is inimical to that goal, and that the goal is desirable, they may still impose the penalty non-intentionally. Their only aim in taking their custom elsewhere may be to get a better bargain: they may have no wish to help police her into competitive pricing.

The third and last distinction necessary for our taxonomy marks a divide among decentralised, non-intentional sanctions. We describe it as the distinction between contingently and essentially non-intentional sanctions. The non-intentional sanctions imposed by consumers on the non-competitive vendor are contingently non-intentional. The consumers intentionally go elsewhere for a better price, and, while it is not intentional on their part that doing so punishes the original vendor for non-competitive pricing, they might also have performed the action intentionally under that description; they might also have wanted to show the vendor, as we say, that she could not get away with that sort of behaviour.

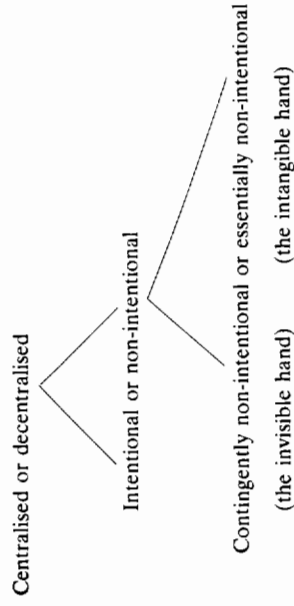
But a non-intentional sanctioning response does not have to consist in

an action that just happens to be non-intentional under the sanctioning description. It does not have to be just contingently non-intentional. The point is not often noticed but a little reflection makes clear that the sanctioning response may not consist in an action at all: it may consist in the formation of an attitude, for example, in the response of thinking badly or thinking well of the agent in question. I am punished by being observed looking through the keyhole of a hotel bedroom, even if the observer does nothing: even if the observer wishes he were a hundred miles away (Pettit 1990). That I am punished is clear from the way in which I smart with embarrassment, when I realise there is someone watching me and, no doubt, thinking ill of me. The sanctioning imposed by the other in such a case is clearly non-intentional. And clearly it is essentially non-intentional, for while the observer might have intentionally stationed himself in the corridor, and might have intentionally kept me under observation, he could not have come to think ill of me intentionally. Thinking well or ill of someone, like thinking that it will rain tomorrow, is not the sort of thing one can ever do intentionally. It is an essentially non-intentional response.

With these distinctions in play we can formulate the two hypotheses of this paper. The first hypothesis is that invisible hands are ordering mechanisms which operate, at least mainly, via decentralised and *contingently* non-intentional responses. The second is that there are also ordering devices which operate via decentralised and *essentially* non-intentional responses; these devices we describe as intangible hands. The next section is given to a discussion of the invisible hands hypothesis, and the third section to a discussion of the hypothesis about intangible hands.

But before leaving this section it may be useful to map the distinctions made, and the hypotheses formulated, on a tree diagram.

The Nature of Controls



2. INVISIBLE HANDWORK

2.1. *The Central Examples*

Adam Smith introduced the term 'invisible hand', in the sense in which it is now used, in this passage from *The Theory of Moral Sentiments*, first published in 1759.

It is no purpose that the proud and unfeeling landlord views his extensive fields, and without a thought for the wants of his brethren, in imagination consumes himself the whole harvest that grows upon them. The homely and vulgar proverb, that the eye is larger than the belly, never was more fully verified than with regard to him. The capacity of his stomach bears no proportion to the immensity of his desires, and will receive no more than that of the meanest peasant. The rest he is obliged to distribute among those, who prepare, in the nicest manner, that little which he himself makes use of, among those who fit up the palace in which this little is to be consumed, among those who provide and keep in order all the different baubles and trinkets, which are employed in the economy of greatness; all of whom thus derive from his luxury and caprice, that share of the necessities of life, which they would in vain have expected from his humanity or his justice. The produce of the soil maintains at all times nearly that number of inhabitants which it is capable of maintaining. The rich only select from the heap what is most precious and agreeable. They consume little more than the poor, and in spite of their natural selfishness and rapacity, though they mean only their own convenience, though the sole end which they propose from the labours of all the thousands whom they employ, be the gratification of their own vain and insatiable desires, they divide with the poor the produce of all their improvements. They are led by an invisible hand to make nearly the same distribution of the necessities of life, which would have been made, had the earth been divided into equal proportions 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. (Smith 1982, pp. 184–85)

Do the rich really 'advance the interest of the society' in this way? Do they really ensure a more or less equal distribution of the necessities of life, as Smith alleges? The point is arguable. But our interest is in the nature of the claim made in the passage, not in its accuracy. The passage presents the original example, as we might call it, of an invisible hand mechanism, and the question is whether it gives support to our thesis about the characterisation of that sort of mechanism. Our thesis, as advertised earlier, is that the invisible hand is characterised by relying mainly on controls that are decentralised and contingently non-inten-

tional. Assume then that the distributive mechanism described by Smith really operates as he says it does. The question is whether it operates by means of the sort of controls that our thesis postulates.

There are two sorts of controls invoked by Smith. The main one is the positive sanctions available to the rich if they will only part with the necessities of life that are surplus to their requirements: in particular, if they will part with them in return for the services of others. The rewards available to the rich are precisely those services, services which enable them to enjoy baubles and trinkets, luxury and caprice. The sanctions are decentralised, for they are conferred severally by those with whom the rich deal. The sanctions are non-intentional, for in all likelihood the sanctions provided by the poor are not provided in order to reward the rich for parting with their surplus; they are provided in order to earn a living. And the sanctions, finally, are contingently non-intentional, for, while it is not actually intentional on the part of the poor that they reward the rich in this way, they just might have provided the rewards intentionally; they might have individually done so, for example, out of a religious desire to get the rich to do what is seen as the right thing.

These comments are enough to show that our account of the invisible hand is vindicated in the case of Smith's example, for the distributive mechanism he describes relies mainly on the sanctions discussed, and these are, as we have seen, decentralised and contingently non-intentional. But there is a second sort of control invoked implicitly in Smith's discussion and it is also worth drawing attention to this. Smith asserts that at all times the soil maintains just about the number that it is capable of maintaining; there is neither over-population nor under-population, as we might say. The proposition that there is no over-population is essential to his theory, since the rich will not have enough necessities to distribute to all, or to distribute to all at an appropriate level, if there are more people than the soil can maintain. But what control is going to ensure that the population is at a suitable level? Here it seems that Smith is going to have to rely on a screening device. He is going to have to say that there will not ever be long-term over-population, because the under-nourishment it entails will soon reduce the number of those around; it will screen out the 'surplus population'.

How does this screening control fit into our taxonomy? The classification of the distributive mechanism as an invisible hand depends only, by our account, on the nature of the controls that are mainly involved: in this case, the sanctions on the rich. But it is interesting to note that

the screening device invoked is also decentralised and contingently non-intentional in character. It is decentralised, so far as it does not operate via a central screening agency or anything of the kind. It is non-intentional, so far as those who fail to share with the needy do not do so with the intention of keeping the population at a certain level. And it is contingently non-intentional to the extent that, still, those who fail to share might have refused to do so out of that desire.

Smith's example of the invisible hand bears out our characterisation. An invisible hand is a device that brings about a presumed good via the operation of decentralised and contingently non-intentional controls. Why call the device invisible? Presumably, because those who remain mere participants in the system, those who fail to adopt a theoretical stance on what happens, will necessarily fail to recognise what is going on. It is definitional of not having attained a theoretical perspective on the system of distribution – of not having attained Smith's insights, in particular – that one does not recognise the operation or the effect of the sanctions in question. Participants who are not also theorists are embedded in their individual positions, and are aware of the immediate pushes and pulls that work on them; but they lack any sense of the aggregate shape of things. The hand that moulds things to that shape remains invisible within their perspective. Smith stresses the invisibility of this ordering device to the rich: 'without intending it, without knowing it, (they) advance the interest of the society'.² But the rhetoric makes that emphasis the natural one, and it is clear that he also takes the poor to be in ignorance of the device that benefits them.

If the founding example of the invisible hand bears out our first thesis, that the invisible hand is characterised by relying mainly on decentralised and contingently non-intentional controls, so do other examples that are more or less like the founding case. Thus consider the example offered by bargaining in general. When people enter into bargains with one another, and when there are no relevant negative external effects, those bargains have the benefit of contributing to a Pareto-improvement: some people have their preferences satisfied; none have their preferences frustrated. Thus people produce an overall Pareto-improvement by rewarding one another's bargaining activities through entering into such bargains. This mutual rewarding and reinforcement is a decentralised form of sanctioning. And it is also a contingently non-intentional sanctioning exercise. The parties do not

reward one another in order to produce the Pareto-improvement, only in order to better their own individual fortunes. And this is a contingent fact about them, for, however unlikely, it is abstractly possible that they might have rewarded one another out of a desire for such an overall improvement.

The two cases considered so far deal with economic agents. A further, fairly traditional example of the invisible hand involves political agents. Consider politicians who rely on electoral support for continuing in office and who have to respond to lobbying pressures from various interest-groups while they are in office. Suppose, as various traditions would suggest, that the lobbying responses of politicians should be fair in relation to the different interest-groups in society: that political decisions should reflect the extent and intensity of the different interests involved. Liberal, pluralist theory proposes that under democratic conditions we can rely on an invisible hand process to ensure that that effect is promoted. The idea is that the different interests will each have their lobby group, that politicians will try to respond to lobbying pressures in a way that does most for the prospect of their re-election, and that if they do this then they will take appropriate account of the intensity and extent of the interests in question: this, on the assumption that they believe that voters will vote according to those interests (but see Brennan and Pettit 1990). The mechanism invoked here is traditionally presented as an invisible hand; and so it ought to be, on our analysis. The relevant sanctions on the politicians are decentralised, because it is the voters severally who reward them. And the sanctions are contingently non-intentional. The voters might have rewarded the politicians out of a pluralist desire for some balance in the representation of interests, but it is as certain as these things ever get to be that in the actual world voters do not reward them out of that sort of desire; they impose the rewards non-intentionally.

2.2. *Three Further Examples*

The three examples already considered are all fairly straightforward cases of invisible hand mechanisms. But the analysis extends naturally to more complicated cases, too. We shall try to make the point with some brief comments on three such instances: the competitive market; conditional cooperation; and conventional coordination.

In the competitive market producers are drawn, so it is alleged, to

sell at the competitive price – the price which equals marginal cost of production – and thereby generate an important social benefit. The assumption is that producers are each self-interested and rational and that in any area of production new producers are free to enter at any time. This means that in each area of production everyone will be better off if all cooperate in fixing a price than if all defect, but that defection will always be better for each than sticking with any agreed non-competitive price: defection will promise to increase that individual's share of the market more than is sufficient to compensate for the lower price he charges. Each area of production will involve a free-rider problem and as the individual purchasers seek to free ride on the fidelity of others to any agreed price, they will force the price down to the lowest at which they can stay in business. The idea is that in a competitive market the price of every good produced will be forced down to marginal cost and everyone will share in that assumed public benefit.

Now the case of conditional cooperation. Imagine two people who are each better off if each gives good information in answer to the queries of the other than if neither does so: mutual truth-telling, say on the location of food, is a Pareto-superior outcome. But each finds lying dominant in this sense: if other things are equal, it is better for each to lie, no matter what the other does. Suppose further, surely with some plausibility, that the parties are involved in an indefinitely iterated series of information-seeking encounters, now with one party, now the other, being in possession of the desired facts. The literature on conditional cooperation has shown that this mechanism ensures, under fairly plausible assumptions, that other things will not be equal, at least when only two parties are involved (see Pettit 1986), and that universal truth-telling may still emerge and stabilise (see Taylor 1976, 1987b; Hardin 1982; Axelrod 1984; Sugden 1986). Over the series of encounters the parties have to adopt a strategy like permanent cooperation or defection or a conditional strategy like tit-for-tat: that is, cooperate in the first round and then do whatever the other did in the previous round. The literature argues that a conditional strategy like tit-for-tat may well emerge in such a situation: it represents an equilibrium that is Pareto-superior to permanent defection, being better for some – in fact for all – and worse for none. If it does emerge, then it will secure the presumed good of universal truth-telling.

The third case, that of conventional cooperation, has been definitively

analysed in Lewis (1969). Suppose that in a certain population people regularly confront one another in coordination predicaments involving pairs or other small groups: they drive cars and meet on narrow roads where it is important that they each go left or each go right; their telephone conversations are regularly cut off and it is important that the original caller or the person called tries to call back, not both of them; and so on for similar cases. Lewis has shown that the reward of doing the same as the other person – the reward of driving on the same side of the road or leaving the task of calling back to the same party – should ensure in a variety of circumstances where one course of action is the salient candidate that the individuals involved will converge on that behaviour. But what is salient in one situation will tend to be salient in others, and so the argument also supports the claim that corresponding conventions of behaviour – a presumed public good – will emerge and stabilise across the relevant group. Lewis argues that such regularities will be conventions in that: almost all individuals conform, almost all expect others to conform, and nearly everyone prefers to conform if others conform; and, moreover, it is common knowledge in the group that these conditions are fulfilled.

In all of these more complex cases it should be clear, as in the more straightforward ones, that people produce a presumed good under the influence of decentralised and contingently non-intentional sanctions. The sanctions are decentralised so far as they are imposed severally by other agents in all three cases. And case-by-case analysis shows that the sanctions are also all contingently non-intentional. Consumers might intentionally punish vendors for selling above competitors, say, because of wanting to promote competitive pricing. The parties in the tit-for-tat case might punish any defection with the intention of bringing about universal truth-telling. And people in the conventional exchanges might make things difficult for those who do not suitably coordinate their behaviour, with the intention of promoting a certain convention. But in none of these cases is it plausible to think that the sanctioning is pursued in such an intentional manner. More likely, the consumer will be looking only to her own immediate advantage, the tit-for-tat to getting her own back on the defector, and the person in the conventional exchange to achieving the best result for herself. The sanctioning implicit in their behaviour is contingently non-intentional.

Why do we distinguish these three more complex cases from the earlier examples? They differ from those earlier examples in an interest-

ing way. Suppose that an action-type is sanctioned – say, rewarded – in the decentralised, non-intentional fashion of the invisible hand. The consequent pursuit of that action-type may directly produce the good associated with the invisible hand, as more and more agents make their individual contribution: as more and more rich people seek out the services of the poor; as more and more parties to the market make satisfactory bargains; and as public officials respond appropriately to more and more lobby groups. But the pursuit of the action-type may also produce that good in an indirect fashion, a fashion that requires agents to be aware of one another's behaviour and to adjust to one another in how they pursue the action-type. That is what happens in the three complex cases. Thus the producers who seek not to be undercut by competitors do not produce competitive pricing by an aggregation of individual efforts; rather, they produce it by responding to the efforts of others not to be undercut, until the price is driven down to the competitive level: to the level where it equals marginal cost. In the complex cases, the pursuit of the appropriate action-type is beneficial only so far as the pursuit is interactive and dynamic.

We have discussed six mechanisms that have a reasonable claim to be regarded as invisible hand devices: three fairly straightforward, three of a more complex kind. In each case we find that the mechanisms operate mainly via controls – usually just sanctions – that are decentralised and contingently non-intentional. This is good evidence in favour of our first thesis, that the best way to characterise invisible hand mechanisms is by the fact that they rely principally on controls of that sort. We take our stand on that evidence and place our proposed analysis of invisible hands at the mercy of whatever counterexamples can be produced.

2.3. *The Invisible Backhand*

But all the cases we have considered so far involve the production of an aggregate effect that is, so we assume, a social good; hence we talk of an invisible hand rather than an invisible backhand. Before leaving this section, it may be useful to take a quick glance at the backhand. The most prominent general case of this phenomenon is the unsolved collective action problem: most notably, the n-person prisoners' dilemma. Here, the structure of penalties and/or rewards embodied in the ordering mechanism encourages actors to choose those options that

produce a socially undesirable outcome – say a Pareto-inferior position. Examples are so familiar from the 'market failure' literature that we need merely list them: the littered beach; the congested road; tax avoidance; shirking within the large team; the rat race; and so on. In all such cases, the aggregatively damaging behaviour is produced, at least in part, by the fact that agents let one another's cooperative efforts go unrewarded; indeed they let one another suffer the sucker's penalty of making a cooperative effort that yields no return, if they go it alone. Such penalisation is decentralised, being practised by the agents on one another. And it is also contingently non-intentional: the agents are assumed to impose these penalties on one another, not with a view to punishment for cooperation, but in the pursuit of their own special goals – usually their own advantage.

For a more specialised example of an invisible backhand, consider T. C. Schelling's story as to how total residential segregation – something we may take for our purposes to be socially undesirable – can come about, whether between blacks and whites, Catholics and Protestants, or whatever group (Schelling 1969). Each party in a mixed neighbourhood has eight neighbours, as a square on a draughtsboard has at most eight adjacent squares. Suppose that each comes to be dissatisfied with a situation where there are fewer than four of his kind among his immediate neighbours. A little experiment on a draughtsboard shows that if each party assumes this not particularly bigoted motivation, then, as by an invisible backhand, a short sequence of moves will lead to total segregation. Those parties in unsatisfactory situations will move, and, by moving, they will make the situations of other parties unsatisfactory – they will penalise other parties – up to the point where there is nearly total segregation. The sanctioning that produces this unhappy state of affairs is decentralised and contingently non-intentional, and so the mechanism at work can fairly be described as an invisible backhand.

3. INTANGIBLE HANDIWORK

3.1. *The Possibility of the Intangible Hand*

Under the invisible hand or backhand, agents police one another into the production of a certain sort of behaviour, by intentionally responding to each other in certain ways. In particular, they do this without it happening to be intentional on their part that the responses have a

policing effect. The mechanism relies on a decentralised, contingently non-intentional pattern of sanctioning. We turn now to the possibility of a different sort of decentralised sanctioning: one which operates, at least in the main, via an attitudinal and therefore essentially non-intentional sort of sanctioning. We are interested in the possibility of a sanctioning system that depends only on the formation of attitudes of approval and disapproval towards certain forms of behaviour, among the relevant population. That is, we are interested in the possibility of a system that exploits our natural, more or less approbative or disapprobative dispositions and puts them into effect in generating certain aggregative results. We describe the sort of device envisaged as an intangible hand.

A word on what it means to approve and disapprove. We intend these terms in a very general sense. Thus to approve of someone may be to feel acceptance, admiration, affection, or gratitude, or, more intellectually, it may be to think well of them, whether in absolute or comparative terms, and whether on a moral or on some other basis. In general it amounts to giving the person a degree of positive regard. Of course we do not approve, or disapprove, only of persons. We also approve, and disapprove, of actions. Such action-directed regard can be paraphrased as regard for the agent on account of her having acted in the manner in question. And similarly for regard given, not to someone's actions, but to their motives, or traits, or goals, or whatever.

The possibility of an approval-centred system of decentralised sanctioning may seem to fall foul of a line of thought often canvassed in the literature. If a certain aggregative result – say, a desirable result – is to be supported by approbative sanctions, as under the intangible hand, then it must be that other people offer approval for an individual's contributing to the result and/or offer disapproval for his not doing so. But now we confront a problem. If people are not sufficiently motivated in general to contribute to the result themselves, how can we expect them to be sufficiently motivated to go around approving of contributions by others and/or disapproving of failures to contribute? It is not obvious, so it will be said, that imposing approbative sanctions on others in this way is going to be less onerous than actually making one's own contribution. And so if we think that people are not sufficiently motivated to make that contribution, then we must have doubts about whether they are likely to bother with approving of others for making it or to bother with disapproving of them for not doing so.

This problem is sometimes raised in the form of a paradox. Assume, as will often happen, that people's contributing to the result amounts to their overcoming a collective action predicament of some kind: say, it amounts to their solving a free-rider problem, with each making a suitable contribution rather than relying on others to make the contributions required. Suppose it is suggested that people may each be induced to act appropriately, with the result that the predicament is resolved, so far as they each approve of such action on the part of others, and disapprove of inaction. How can that suggestion survive the objection that if there was a collective action predicament raised by the desired behaviour, there will be a similar predicament with the relevant approbative sanctioning? If people were motivated to free ride on the behaviour of others in the original case, won't they be motivated to free ride on the sanctioning efforts of others in this one? The suggestion as to how the original predicament might be resolved seems to assume the spontaneous resolution of a derived predicament – the enforcement dilemma, as it is sometimes called – which is just as problematic as the first (Taylor 1987b, p. 30).

There are a number of queries to raise against this objection. For example, it assumes that approbative sanctioning of others generally involves costs – say, that it takes the form of rebuking offenders, thereby suffering embarrassment and inducing resentment – whereas sanctioning may often be rewarding, as in praise or flattery. But such more or less detailed queries apart, there is a fundamental misconception at work in the objection. This is the assumption that when an individual agent is approbatively rewarded or punished by others, then the others must intentionally do something if the individual is to be sanctioned; the others must make an effort of some kind, rousing themselves to the task of approbative reward or punishment: they must do or say something that affects the agent in a suitably reinforcing or inhibiting way. But the sort of assumption at work here is mistaken.³ People can sanction an agent without making any intentional effort whatsoever. They can sanction him just by the fact of involuntarily forming certain attitudes of approval or disapproval, and being assumed to form such attitudes, in response to what he does (Pettit 1990; forthcoming, Ch. 6). More dramatically indeed, they can sanction him just by the fact of being assumed, perhaps wrongly, to form such attitudes in response to his action.

James Buchanan mentions two costs that he thinks are associated

with approbative sanctioning. "Enforcement has two components. First, violations must be discovered and violators identified. Second, punishment must be imposed on violators. Both components involve costs" (Buchanan 1975, pp. 132–33). But people may not have to do anything to identify a violator: they may just be around in sufficient numbers to make it highly likely that some of them will notice, and be assumed to notice, the doer and the deed. And equally they may not have to do anything to make the agent suffer or rejoice in such a case. The fact that they are rightly or wrongly assumed to notice the agent, under circumstances where his behaviour is rightly or wrongly assumed to attract their unfavourable or favourable regard, may be penalty or reward in itself.

Imagine that someone comes upon you looking, as in the example given earlier, through the keyhole of a hotel bedroom; or imagine that someone stumbles on the fact that you have made a large contribution to some charity. In the first case you will probably smart under the gaze of the other, and in the second you may bask – assuming embarrassment allows – even if the other party remains inactive and unexpressed. It is common knowledge in our society that the first sort of act will tend to occasion an unfavourable regard, the second a favourable one, and being regarded favourably or unfavourably is a motivating property in itself: it counts for or against the action in question, though it may not be a decisive consideration either way.

As against Buchanan, then, we argue that people can reward and penalise in a decentralised way without having to make any effort. They may reward or penalise a sort of behaviour just by being in a position to observe, and be assumed to observe, it. The sanctioned agents are sanctioned through knowing or thinking they know what the observers must think. There is no necessary requirement that those observers actually give expression to their thoughts. As Adam Smith puts it:

We are pleased to think that we have rendered ourselves the natural objects of approbation, though no approbation should ever actually be bestowed upon us; and we are mortified to reflect that we have justly merited the blame of those we live with, though that sentiment should never actually be exerted against us. (Smith 1982, p. 116)

Thinking of electric fences and gun traps, Buchanan (1975, p. 131) observes that "we need not reach into the extremities of science fiction to think of devices that could serve as automatically programmed en-

forcers". We agree, though with rather a different thought in mind. The upshot of our approach is that people are automatic, involuntary sanctioners of one another's behaviour. They confer certain rewards and penalties on one another without exacting any price or making any contract, because they confer those rewards and penalties in an essentially non-intentional way (Pettit 1990).

Intangible hand devices, as we envisage the category, rely on purely attitudinal responses – rightly or wrongly imputed – as controls for producing the associated results. The attitudinal responses in play lack the palpable quality of actions and that is why we describe the devices as intangible. We should now be getting on to the illustration of intangible hand devices, but, before we do so, it will be useful to say something on the motivational assumption that supports the idea that there are such devices to be found.

We did not say much on the motivational assumptions associated with the invisible hand, though there were some. We assumed, to recall examples, that the rich have desires that the poor can satisfy, politicians have desires that voters can satisfy, business people have desires that consumers can satisfy, and the like. We assumed, as it is sometimes put, that people have a desire for economic gain. In order to motivate the idea of the intangible hand, we have to be explicit about one further motivational assumption. This is that people generally have a desire for social acceptance as well as economic gain, whether social acceptance is desired intrinsically or in virtue of promising economic gain. More specifically, they have a desire to be regarded favourably rather than unfavourably: to be regarded, say, with admiration, gratitude, or pleasure. The property of an option that it tends to produce favourable regard makes for its desirability in the eyes of an agent, even if in many cases other properties trump it and render the option undesirable.

This assumption ought to be overwhelmingly plausible. It is going to be rational in terms of economic gain for any agent to desire to be favourably regarded, since, other things equal, the more others regard him favourably – and the more favourably they regard him – the more likely they are to treat him with favour in the future: say, on occasions when they are in a position, without cost to themselves, either to benefit him or to benefit someone else. To be regarded favourably is to enjoy the benefit of an increased probability of favourable treatment; to be regarded unfavourably is to suffer a reduced probability of such treatment.

Suppose it is rational in terms of economic gain for an agent to desire something X, as we say it is rational to desire social acceptance. If the agent actually desires X, that may be fixed in any of a number of ways: because he actually recognises the economic value of X; because he is in the habit of desiring X, whether for reasons of past recognition of its economic value or of childhood training; or because evolution has preempted his individual reasoning and implanted in him a non-instrumental desire for X, a desire for X as for an intrinsic good. We shall assume that people desire social acceptance without generally recognising its economic value in an explicit way, though our argument would go through – most of the examples given later would be vindicated – even if the desire for favourable regard was always mediated by an economic estimate of its value. The important point in the argument is that an agent can be sanctioned at a given time without the sanctioning agents actually doing anything then: they need only be in a position such that the agent rightly or wrongly assumes – or would assume, if he behaved inappropriately – that they will have come to think the better or the worse of him for what he has done.

The assumption that as a matter of fact people desire social acceptance or favourable regard without always making an estimate of its economic value will make the presentation of our case easier. It is surely reasonable, since introspective and empirical evidence suggests that the desire for favourable regard is not parasitic upon an appreciation of any derived economic benefit. Consider how you smart and flinch beneath the gaze even of strangers, when you realise that you have been noticed doing something that is culturally compromising or demeaning: it may be as humdrum an activity as browsing over another person's mail, breaking a red light, or just picking your nose. And consider how often people care for what will be thought of them after death – even when those who think ill cannot inflict any relevant harm: they cannot hurt the bereaved family or whatever.

The postulate we are supporting, even in this strong form, has the authority of a long tradition. Here are some illustrative quotations.⁴

Cicero: "The sages, taking nature as their guide, make virtue their aim; on the other hand, men who are not perfect and yet are endowed with superior minds are often incited by glory, which has the appearance and likeness of *honestas*".

Aquinas: "Honour is the greatest good among external goods".⁵

John Ray (1713): "I cannot but admire the Wisdom and Goodness of God, in implanting

such a Passion in the Nature of Man, as Shame, to no other Use or Purpose, that I can imagine, than to restrain him from vicious and shameful actions".

Christian Wolff (1732): "Nothing pre-eminently great has ever been done in the world which did not flow from glory as its source".

Voltaire (1734): "It is pride above all that has been the principal instrument with which the fair edifice of society has been built".

Hume (1739): "There are few persons, that are satisfied with their own character, or genius, or fortune, who are not desirous of shewing themselves to the world, and of acquiring the love and approbation of mankind".

Adam Smith (1759): "Nature, when she formed man for society, endowed him with an original desire to please, and an original aversion to offend his brethren. She taught him to feel pleasure in their favourability, and pain in their unfavourable regard. She rendered their approbation most flattering and most agreeable to him for its own sake; and their disapprobation most mortifying and most offensive". (Smith 1982, p. 116)

Kant (1786): "A craving to inspire in others esteem for ourselves, through good behavior (repression of that which could arouse in them a poor opinion of us) is the real basis of all true sociality".

The motivational assumption on which we are relying, even in its strong form, should also commend itself to current ways of thinking. It is obviously going to be congenial to those in the sociological tradition of social theory. Within that tradition the desire for status ranks with the desire for wealth and power – as it does in Kant's *Anthropology* – as one of the basic human motives. To enjoy status is to enjoy a special kind of acceptance by people at large: specifically, a greater acceptance than relevant others. But the assumption ought also to recommend itself to thinkers who align themselves with the economic or rational choice tradition. John Harsanyi offers the following postulate as a principle of explanation for rational choice theory. "People's behavior can be largely explained in terms of two dominant interests: economic gain and social acceptance" (as quoted in Taylor 1987a, p. 66).

The desire for the regard of others may not be the most admirable component in a psychology, but it enables us to see how there can be decentralised sanctions which are essentially non-intentional, unlike the sanctions associated with the invisible hand. It explains how a property which is not sufficiently motivating for agents in general – say, they are not selfless enough to try to realise it in their behaviour – can still be a property whose realisation or non-realisation by another goads them to produce a sanctioning response. It explains how agents may police one another into a sort of behaviour which few of them are motivated

to produce without the policing of the others. In short, it explains how the intangible hand is possible. Some such regard-motivation is sufficient for intangible handiwork and, so far as we can see, necessary: it represents the only possible source of energy for the hand to draw on.

3.2. *Examples of the Intangible Hand*

It is time to turn to some examples of intangible handiwork. There are three cases at which we shall look. The first involves what we describe as the committee arrangement, the second the incorporation device, and the third the commons mechanism. As we shall emphasise again, we do not say that the intangible hand operates unfailingly in all these instances, only that it seems capable of doing so under certain circumstances. All three cases are putative examples of the intangible hand at work to a purpose that suits the relevant agents, and the society in general, and we shall conclude our discussion with a query as to whether there are also examples available of an intangible backhand.

The first case, for us the paradigm of intangible handiwork, is the committee arrangement. The best example of this is provided by the jury system.⁶ It is important for the well-being of all of us, including jurors themselves, that jurors be conscientious, and known to be conscientious, in assessing the impact of evidence on guilt. But the property of conscientiousness may not be sufficiently motivating for all jurors, and we cannot necessarily rely on its moral attractions to ensure its realisation. Conscientiousness takes time and trouble, and while everyone is better off if every juror is conscientious than if none is, everyone is best off in self-interested terms if he can be the one juror who gets away with being less than conscientious. So how can we expect jurors, as surely many of us do, to be conscientious? We raise the question with regard to the jury, but we might equally have done so with any of a variety of committees. Other suitable examples would be the appointments committee in any firm or bureaucracy, the promotions committee in a university, and the committee for deciding who will be selected for certain honours. Whenever there is such a committee there is a sort of behaviour that is plausibly designated as socially desirable: this we can describe as conscientiousness on the part of the members in their assessment of how candidates do in relation to the brief of the committee. And whenever there is such a committee there is a question as to how we expect members to be conscientious, for that desirable

property may well fail to be individually motivating, even if it is commonly recognised to be suitably desirable.

The committee arrangement puts such decentralised, essentially non-intentional sanctions in play as make conscientiousness more probable than it would otherwise be among the members of a jury; that is why we can reasonably expect jurors to be conscientious. The jury room is hidden from external view, so the jurors ought not to be subject to intimidation. The members of the jury are vetted to ensure that none has a special interest in the outcome, so equally the jurors ought not to be susceptible to bias. The committee arrangement comes into play by ensuring that the fertile motivational ground prepared by those measures is seeded with incentives to be conscientious. The jurors are free to ask one another why they intend to vote guilty or not guilty, and so each faces the penalty of being thought silly for putting up a bad case or the reward of being thought sensible for putting up a good one. The committee arrangement ensures that these sanctions will always be there to make conscientiousness more attractive than it would otherwise have been. Motivated by those sanctions, we may expect that jurors in general will find that the most natural course for them to take is to be conscientious. Either jurors will either engage the bad opinion of others or have to pretend to be conscientious, which must involve just about as much trouble as the real thing. The sanctions in play are decentralised – the jurors sanction one another – and essentially non-intentional – they sanction one another without necessarily doing or saying anything – and so the committee arrangement is an instance of the intangible hand.

Considered in more analytical detail, the committee arrangement serves first to activate the regard-motivation of members of a committee, and second to give it a privileged position. It serves to activate the motivation in three steps: by making conscientiousness a salient good; by making it a presumptive dimension of assessment; and by making it a feasible dimension of assessment. Conscientiousness becomes a salient good, a property whose desirability is manifest to members, because the committee can fulfill its brief only if members are conscientious, and that should be clear to anyone who is introduced to the brief. Conscientiousness becomes a presumptive dimension of assessment, a dimension on which every committee member may expect to be well or ill regarded, so far as the following holds. Every committee member must expect that people who endorse the committee brief –

as the members may generally be presumed to do – would praise conscientiousness *pro tanto*, or censure the lack of it, at least if the costs of doing so were acceptable; that is to say, every member must expect such others to be disposed in suitable circumstances to praise or censure in this manner; and that is to say in turn that every committee member must expect such others to give favourable regard to conscientiousness *pro tanto*, and unfavourable regard to the lack of it. Finally, under the committee arrangement conscientiousness becomes a feasible dimension of assessment, because every member of the committee can be interrogated by the others as to her reasons for wanting to vote one way or another and because the others are therefore in a position to make a secure judgment as to whether she is conscientious.

These aspects of the committee arrangement activate the regard-motivation of members in the sense that they ensure a reward of favourable regard for conscientiousness, a penalty of unfavourable regard for the lack of conscientiousness. Such sanctions may be effective without serving as the actual reasons why members tend to be conscientious rather than not. They may be effective through being standby supports for conscientiousness; through ensuring that were people's actual reasons, whatever they are, to lead them to be unconscientious, then those reasons would soon come to be questioned and reversed. But the fact that the committee arrangement activates sanctions of regard in this way does not mean that the sanctions will necessarily be effective. The arrangement may activate regard-motivation without deactivating other, more powerful motives and without having much effect therefore on how things go.

But the committee arrangement does in fact attempt to fulfil this task also, privileging as well as activating regard-motivation. It privileges the motivation, to the extent that it does two things to reduce the impact of other, conflicting desires. First it requires that the members of the committee are selected in such a way as to guard against having a substantial number with a special interest in the outcome and a potential motive therefore to be unconscientious; here it relies on a screening device. The jury is vetted by the defence, and the other sorts of committees are selected according to certain procedures: ideally, procedures that require a certain representational mix and deny any one individual discretion, or at least unchallengeable discretion, in the naming of committee members. And second the committee arrangement requires that the deliberations of the committee be confidential,

so that particular members are less open to being bribed or intimidated by outside parties. These measures are not guaranteed to be successful, but at least they raise the probability that the regard-motivation activated under the committee arrangement will not be trumped by motivation of any other kind; regard-motivation will have a privileged status.

The committee arrangement is the paradigm, in our view, of intangible handiwork. It is an arrangement which stands a very good chance, or so our model suggests, of getting the committee members to generate the desired sort of behaviour. But though we hold up the committee arrangement as a paradigm, we do not mean to suggest that it cannot or does not break down. There are a variety of ways in which the institution may be vitiated and its purposes defeated. The selection and confidentiality procedures may break down, allowing the regard-motivation to be undermined. The committee may be stacked by those with a special interest in having a particular outcome; or the committee may be properly sensitive only to the claims of members of a particular sub-group: say, men but not women; or the effect of the desire for regard, in particular the desire for the regard of other members, may be neutralised by outside intimidation. And at a more basic level the machinery for activating the relevant sort of regard-motivation may fail to work. That machinery, as we saw, normally ensures that conscientiousness is a salient good, a presumptive dimension of assessment and a feasible dimension of assessment. But a widespread cynicism would undermine the salience; an awareness of that cynicism would mean that conscientiousness would cease to be a presumptive dimension of assessment; and the smallest shift in procedures – say, secret voting without discussion – would make it an infeasible dimension of assessment. Besides, the machinery would become utterly useless if the members of committees generally became 'shameless' in relation to one another. A society can become so politically divided that committee members care only for the regard of those in their particular sub-group, be the sub-group one of colour, creed, gender, or whatever; and when that happens there is clearly no hope of getting the committee arrangement to work. The committee presupposes a degree of consensus.

Mention of consensus leads naturally to the second sort of intangible hand device that we would like to mention: this we call the incorporation device or arrangement. Suppose, in line with familiar political realities, that a society involves a variety of interest-groups of which each stands to benefit from making a certain sort of accommodation

with the others: as we might put it, they have common as well as conflicting interests. The lobbying device discussed in the last section is an invisible hand mechanism which is designed to cope with such a situation, encouraging the socially desirable pattern of having the different interests satisfied in proportion to their extent and intensity: this is the device envisaged by a certain sort of liberal theorist, sometimes described as a pluralist. The incorporation device is an intangible hand alternative to the lobbying mechanism, and it is associated with theorists who describe themselves as corporatists rather than pluralists.⁷

The incorporation device presupposes that the groups in conflict recognise their common interest and are willing to be involved or represented in discussions as to how that interest should be promoted or how their differences should be reconciled. The background ensures that in a suitable form of discussion it must be desirable in the eyes of all that the parties involved address themselves conscientiously to the question of how to promote the common interest. Such conscientiousness could fail through sheer lack of effort or insight, but more obviously it could fail through some of the parties putting forward arguments or proposals that ignore the claims of certain of the interest-groups involved. Now the idea with the incorporation device is that as the different parties are incorporated in the effort to promote a common interest, conscientiousness will become a salient good, a presumptive dimension of assessment and a feasible dimension of assessment. The forum will operate, so it is envisaged, like a kind of committee, with the different parties to the discussion assuming a senatorial role as they are induced by the desire for one another's regard to look conscientiously at how sectional interests can best be reconciled. The different parties will be subjected to decentralised, essentially non-intentional sanctions that lead them, as if by an intangible hand, to behave senatorially.

The incorporation device is bound to be less secure than the committee arrangement which it seeks to emulate. The reason is that from the point of view of any representative in the common discussion there will be two groups whose regard matters to her and they may differ in the courses of action they regard favourably. On the one hand, there is the incorporated group of discussants, which gives favourable regard to senatorial responses that seem generally designed to advance the common interest. On the other, there is the non-incorporated group of those whose representative she is, and this group will tend to regard

more favourably responses which give vivid expression to their sectional interest, even responses that are less likely to secure an accommodation with other groups. What guarantee have we that the representative won't opt for the enjoyment of the latter regard, assuming the role of an agent of the sectional interest rather than a senator? If such an agent role is assumed by most, then the forum will cease to be the sort of thing envisaged by proponents of the incorporation device; it will become merely an arena for displaying one's sectional loyalties.

It is a good question in the theory of institutional design as to how we can get the incorporation device to work, if at all. Various measures might be considered. One is shrouding the proceedings of the forum in secrecy, as happens in some countries with bodies on which government, business, labour, and consumer interests are represented. Another is making the representative relation more complex, so that each member of the forum has to answer to a population – say, geographically defined – which involves a mix of different interest-groups. Yet another, or so it might be argued, would involve a party arrangement under which each representative aligns himself with a group whose success requires the support of people facing a variety of interest-groups. Again there is the radical possibility of doing away with representatives altogether and involving every adult directly in the forum. And finally there is the possibility, which we have investigated elsewhere, of introducing voting procedures that favour the electoral success of senators rather than bargainers (Brennan and Pettit 1990).

The committee arrangement and the incorporation device may well be explicitly designed institutional measures. Our third and last case of intangible handiwork comes from an area where there is no explicit design required. We might call it the commons device, after the example which most vividly illustrates it. The commons device operates when conditions like the following come to be fulfilled spontaneously: there is a pattern of behaviour in a certain constituency such that everyone is better off if everyone else displays it; everyone understands that the pattern is beneficial in this way; no one can display or fail to display the pattern without some others being likely to notice; everyone is disposed, and everyone recognises that everyone is disposed, to censure any offender, and to praise any complier, if doing so is not too costly in time, embarrassment, or whatever: that is, everyone regards compliance favourably, deviance unfavourably, and everyone recognises that this is so; and finally, the costs of compliance, the benefits of deviance,

are not so great that the desire for the regard of others is trumped (see Pettit 1990). In such a situation, we can see that regard-motivation is likely to be active in creating sanctions, and probably effective sanctions, for the promotion of the compliant, beneficial behaviour. The sanctions will be decentralised, being imposed severally by those who notice; and they will be essentially non-intentional, involving just the formation or imputed formation of suitable attitudes on the part of the sanctioners. And so the device in operation will count as an instance of the intangible hand.

The example with which we illustrate this case is that of the traditional commons or common land. Imagine it is generally accepted that the land can bear only ten grazing animals per farmer, so that putting more animals out to pasture is overgrazing. In such a situation it is plausible that our conditions would be fulfilled. Everyone is better off, and understands himself to be better off, if everyone else respects the limit of ten animals: it is commonly recognised that respecting that limit is a suitably desirable property. No one can graze at the limit or overgraze without incurring the notice of others. Anyone who notices overgrazing will regard it unfavourably, and this is generally recognised to be so. And finally there are not such great benefits in overgrazing, such great costs in not doing so – not at least if others comply – that the desire for the regard of others is trumped. We might reasonably expect a pattern of moderate grazing to emerge and stabilise under the sanctions created by the desire for regard.

The example is distinctive, if outdated, but the commons device illustrated may work in a variety of situations. The device can operate to resolve a range of free-rider predicaments of a kind with the predicament of the commons. It may operate, for example, to inhibit people from littering the local park, from taking young fish from the river or sea, or from using their lawnmowers too late at night or too early in the morning. The commons device represents a less intricate form of the intangible hand than the committee or incorporation arrangements, but it is a form in which that hand is probably more frequently in evidence.

3.3. *The Intangible Backhand*

We hope this will suffice by way of illustration of the intangible hand. In conclusion, we would like to turn to backhand cases, as we did in

concluding the discussion of the invisible hand. Is there an intangible backhand? Are there cases where a sort of behaviour that is undesirable is triggered by decentralised and essentially non-intentional sanctions?

If there is such a thing as an intangible backhand, then the following conditions must be fulfilled. First, the backhand must have a socially undesirable impact. Second, it must operate via decentralised sanctions. And, third, it must operate via approbative, attitudinal sanctions. How could the third condition be fulfilled? How could people approve or be assumed to approve of actions which help to produce a social bad, for example, or disapprove of actions which help to avoid it? We must assume, as we have done up to now, that there is no controversy over whether a given bad is bad, or a given good good. Thus we cannot answer the question raised by postulating differences of opinion on such evaluative matters. How then can people's attitudes or assumed attitudes elicit from one another behaviour which produces a result that all regard as bad?

One possibility leaps to mind. This is where some type of behaviour is regarded as individually admirable though its collective realisation leads, more or less surprisingly, to a result that all regret. Consider one sort of thing that can happen on a beach when someone gets into trouble in the water. Many people, spurred or reinforced by the prospect of approval from others, swim towards the drowning person. But so many do this that they get in one another's way, and fail to save the swimmer in difficulty. An intangible backhand operates in such a case, as people's imputed attitudes of approval elicit a form of behaviour among a number of individuals which, collectively pursued, is undesirable by everyone's lights. The sanctioning which secures the result is decentralised and, being attitudinal, is essentially non-intentional.

There are any number of cases where an intangible backhand of this kind can strike. Here is one sort of case that has already received attention in the literature. It is commendable that parents should be concerned about the welfare of their children, or so our common-sense morality has it, and therefore parents may be supported in that special concern by the reward of being well regarded. But if all parents do this, then the children may be worse off than if parents concern themselves with the welfare of children generally (Parfit 1984). The children might, do better, for example, through the collective provision of education than if parents compete to provide an educational advantage for their own children (Hollis 1982). If this is so, then the behaviour for

which people are individually well regarded is such that, collectively pursued, it leads to a result that is bad by everyone's lights. There is an intangible backhand at work which parallels closely the sort of thing that operates in our beach example.

There is a second, more complex way in which an intangible backhand may strike in a community. The best way to exemplify it may be to look at another sort of thing that can happen in the beach case. One of the least approved of activities, paradoxically, is the pursuit of approval itself. "The general axiom in this domain", as Jon Elster (1983, p. 66) puts it, "is that nothing is so unimpressive as behaviour designed to impress". Imagine that it is a matter of common assumption among those on the beach that everyone thinks that swimming to the rescue of the person in difficulty will generally earn approval. The fact that this is commonly believed may lead to its becoming also a matter of common assumption that anyone who does swim, especially in a situation where there are many possible rescuers, does so in order, at least in part, to earn approval. But if that becomes a matter of common assumption, and if no one wants to be thought to be trying to earn approval – no one wants the ignominy of the honour-hunter – then the result may be that no one attempts a rescue, or at least no one rushes to do so. As in the other sort of case an intangible hand may strike, with a potentially fatal effect.

Although this second sort of example of intangible handiwork is complex, presupposing complicated beliefs on the part of the people involved, we think it may be exemplified fairly commonly in ordinary life. We can all envisage the case – the Kitty Genovese type of case, for example – where someone in public view needs help; where anyone who holds back may contribute to the evil of letting the person suffer or die; and where the assumption that everyone recognises this, and that everyone believes therefore that there is honour to be gained from helping, raises the probability that no one will help. No one helps, because no one wants the dishonour of being thought to be an honour-hunter.

CONCLUSION

Our object in this paper has been to defend two propositions: that invisible hands can be usefully characterised in terms of the decentralised and contingently non-intentional nature of the controls they de-

ploy; and that an important complementary device – the intangible hand – is characterised by its use of controls that are equally decentralised but are essentially, rather than contingently, non-intentional. In conclusion, we would like to float a hypothesis on the history of invisible and intangible hands. We are not the ones to judge whether the hypothesis is sound, but we hope we may attract historical scholars to comment.

The invisible hand idea was first properly articulated by the eighteenth-century satirist Bernard Mandeville, who had an important influence on Adam Smith (Hayek 1966). The Mandevillian idea, outlined in his *Fable of the Bees*, was that the individual pursuit of private interest often did better by the common weal than the individual pursuit of public virtue. Arguably, Smith's idea of the invisible hand is essentially that same theme, stripped of the mockery with which Mandeville canvassed it.

The butt of Mandeville's satire was the republican or civic humanist tradition, which, by many accounts, shaped the language of Western political thought from about 1400 to 1800: from the time of the Florentine and the other Italian republics to the founding of the United States (Pocock 1975; Skinner 1983, 1984). This tradition, which goes back to ideals of the Roman republic, in particular to the work of Cicero, stresses the need for public virtue if people are to be able to enjoy liberty securely. In mockery of that theme, Mandeville used the invisible hand idea to argue, in his distinctive tones, that private vice is public virtue.

Was Mandeville's satire justified? In some cases, it may have been. But in general it would be a mistake to think of the republican tradition as having required, in a moralistic and eminently satirisable way, that people must become figures of public virtue if the common weal is to be realised. For what the tradition mostly supposed was that people would never come to be figures of virtue in this sense, and that it was necessary to design institutions to keep them on the virtuous path: to make them behave, particularly when holding public office, as if they were virtuous. Thus Quentin Skinner (1983, p. 11) writes in exposition of Machiavelli on the need for virtue: "By the force of law the people were liberated from the natural consequences of their own *corruzione* and transformed in effect into *virtuosi* citizens".

If we look at the republican tradition of thought, with an eye to identifying the institutions designed for the promotion of public virtue,

what we actually find is very striking. We find that writers in that tradition place great emphasis on the desire for regard, in particular the desire for honour, and that they look for institutions that will exploit this desire and thereby get people to behave virtuously. In particular, such writers look to decentralised, intangible hand devices in trying to see how the public weal might be promoted (Oldfield 1990). Many of the institutional devices which come down to us, at least in conception, from the era when republican ideas were dominant involve an intangible hand component. We are thinking of initiatives like the limitation on tenure of public office; the answerability of officials to a legislative forum; the impeachability of those in public office; the requirement of open debate in the legislative forum; the jury system; the possibility of appeal against judicial sentence; and a system of conferring public honours on those who serve the polity well.

Now to our hypothesis. There are powerful republican themes in the work of Adam Smith, as Donald Winch (1978) has shown, and there is a recurrent emphasis in his work on the desire for regard, as we have ourselves indicated. But we conjecture that the association between the idea of the invisible hand and the Mandevillean attack on republicanism may have led those who followed Smith to neglect earlier republican themes, and in particular the intangible hand idea. The attention and adulation given to the invisible hand, combined with the assumption that it represented a radically new way of thinking, may have blinded people to the more familiar idea of basing institutional order on the working of the intangible hand. That idea may have been treated as if it were of a piece with the moralistic notion of relying on the presence of spontaneous virtue for the orderly working of society.

We think that, whatever its sources, the neglect of the intangible hand since the time of Adam Smith has been unfortunate and unnecessary. For while invisible hands offer important decentralised resources for the ordering of society, they are not the only alternatives to direct, centralised control. Intangible hands represent complementary, not competitive, devices, and they ought to attract as much attention as their better-known counterparts. In the actual history of ideas, the articulation of the one device may have silenced the claims of the other. But we are not condemned to live indefinitely with this silence. We can give a fresh hearing to the claims of the intangible hand without removing the invisible hand from its established place of honour.⁸

NOTES

¹ Perhaps the only connection between the two sorts of spontaneity is that the fact that an order could have arisen spontaneously is presumptive evidence that it can maintain itself spontaneously; hence, it has been suggested, the common interest in as-if stories of spontaneous generation (Ullmann-Margalit 1978, p. 275).

² Notice, too, that Smith stresses the absence of any intention on the part of the rich to 'advance the interest of the society', whereas what we require, in our definition of the invisible hand, is the absence of any intention on the part of the poor to reward the rich for producing that advancement. Even if the rich had the relevant intention, as we see things, the invisible hand would remain in place: the rich would still be offered decentralised and contingently non-intentional sanctions for advancing the interest of society.

³ While rejecting the idea that every act of sanctioning must have individual costs, we can of course admit that the introduction of a sanctioning system, even a system involving individually costless acts of sanctioning, will have an opportunity cost.

⁴ Except where otherwise stated, the quotations are borrowed, where relevant in his translation, from Lecture V in Lovejoy (1961). The quotations in Lovejoy are taken from: Cicero, *De Finibus* (V, 69); Ray, *Three Physico-Theological Treatises* (1692: quote from 3rd ed., 1713, p. 429); Wolff, *Psychologica Empirica* (1732: quote from 1779 ed., p. 325); Voltaire, *Sermons* (1734: quote from 5th ed., 1765, p. 11); Hume, *Treatise on Human Nature* (Bk. II, Pt. II, Sec. 1, pp. 331–32); Smith, *Theory of Moral Sentiments* (1759, pp. 144–45); and Kant, *Kant's Werke* (Prussian Academy ed., VII, p. 113).

⁵ As stated in the index to Aquinas (1958, p. 156): "*Honor est maximum bonum inter bona exteriora*".

⁶ At least under some idealisation. See Jeffrey Abramson, 'The Jury and Democratic Theory', *Journal of Political Philosophy*, forthcoming.

⁷ Jürgen Habermas has been one of the strongest, recent advocates of the forum. For a characterisation and critiques of his arguments, see Pettit (1982) and Elster (1986).

⁸ We are grateful for comments received when a version of this paper was presented at a conference on institutional design at the Australian National University, in December 1989. We are also grateful for individual comments on the earlier version from Conal Condran, Jon Elster, Paul Finn, Alan Hamlin, Frank Jackson, Peter Menzies, and Quentin Skinner.

REFERENCES

- Aquinas, Thomas: 1958, *Summa Theologica*, Biblioteca de Autores Cristianos, Madrid.
 Axelrod, Robert: 1984, *The Evolution of Cooperation*, Basic Books, New York.
 Braithwaite, John and Philip Pettit: 1990, *Not Just Deserts: A Republican Theory of Criminal Justice*, Oxford University Press, Oxford.
 Braithwaite, John: 1989, *Crime, Shame and Reintegration*, Cambridge University Press, Cambridge.
 Brennan, Geoffrey and Philip Pettit: 1990, 'Unveiling the Vote', *British Journal of Political Science* 20, 311–33.

- Brennan, Geoffrey and Philip Pettit: 1991, 'Modelling and Motivating Academic Performance', *The Australian Universities' Review* 34, 4-9.
- Buchanan, James: 1975, *The Limits of Liberty*, University of Chicago Press, Chicago.
- Elster, Jon: 1983, *Sour Grapes*, Cambridge University Press, Cambridge.
- Elster, Jon: 1986, 'The Market and the Forum', in J. Elster and A. Hyland (eds.), *Foundations of Social Choice Theory*, Cambridge University Press, Cambridge, pp. 103-28.
- Hardin, Russell: 1982, *Collective Action*, Johns Hopkins University Press, Baltimore.
- Hayek, Friedrich: 1966, 'Dr. Bernard Mandeville (Lecture on a Master Mind)', *Proceedings of the British Academy* LII, 125-41.
- Hollis, Martin: 1982, 'Education as a Positional Good', *Journal of Philosophy of Education* 16, 235-44.
- Jackson, Frank and Philip Pettit: 1990, 'Program Explanation: A General Perspective', *Analysis* 50, 107-17.
- Jackson, Frank and Philip Pettit: 1992a, 'In Defence of Explanatory Ecumenism', *Economics and Philosophy* 8, 1-21.
- Jackson, Frank and Philip Pettit: 1992b, 'Structural Explanation and Social Theory', in D. Charles and K. Lennon (eds.), *Reduction, Explanation, and Realism*, Oxford University Press, Oxford.
- Lewis, David: 1969, *Convention*, Harvard University Press, Cambridge, MA.
- Lovejoy, Arthur O.: 1961, *Reflections on Human Nature*, Johns Hopkins Press, Baltimore.
- Nozick, Robert: 1974, *Anarchy, State and Utopia*, Basic Books, New York.
- Oldfield, Adrian: 1990, *Citizenship and Community*, Routledge, London.
- Parfit, Derek: 1984, *Reasons and Persons*, Oxford University Press, Oxford.
- Pettit, Philip: 1982, 'Habermas on Truth and Justice', in G. H. R. Parkinson (ed.), *Marx and Marxism*, Cambridge University Press, Cambridge, pp. 207-280.
- Pettit, Philip: 1986, 'Free Riding and Foul Dealing', *Journal of Philosophy* 83, 361-79.
- Pettit, Philip: 1989, 'The Freedom of the City: A Republican Ideal', in A. Hamlin and P. Pettit (eds.), *The Good Polity*, Blackwell, Oxford, pp. 141-68.
- Pettit, Philip: 1990, 'Virtus Normativa: Rational Choice Perspectives', *Ethics* 100, 725-55.
- Pettit, Philip: 1993, *The Common Mind: An Essay on Psychology, Society and Politics*, Oxford University Press, New York.
- Pocock, J. G. A.: 1975, *The Machiavellian Moment: Florentine Political Theory and the Atlantic Republican Tradition*, Princeton University Press, Princeton.
- Schelling, T. C.: 1969, 'Models of Segregation', *American Economic Review* 59, 488-93.
- Skinner, Quentin: 1983, 'Machiavelli on the Maintenance of Liberty', *Politics* 18, 3-15.
- Skinner, Quentin: 1984, 'The Idea of Negative Liberty', in R. Rorty, J. B. Schneewinds, and Q. Skinner (eds.), *Philosophy in History*, Cambridge University Press, Cambridge, pp. 193-221.
- Smith, Adam: 1982, *The Theory of the Moral Sentiments*, ed. by D. D. Raphael and A. L. Macfie, Liberty Classics, Indianapolis.
- Sugden, Robert: 1986, *The Economics of Rights, Cooperation and Welfare*, Blackwell, Oxford.
- Taylor, Michael: 1976, *Anarchy and Cooperation*, Wiley, London.

Taylor, Michael: 1987a, 'Rationality and Revolutionary Collective Action', in M. Taylor (ed.), *Rationality and Revolution*, Cambridge University Press, Cambridge, pp. 63-97.

Taylor, Michael: 1987b, *The Possibility of Cooperation*, Cambridge University Press, Cambridge.

Ullmann-Margalit, Edna: 1978, 'Invisible-Hand Explanations', *Synthese* 39, 263-91.

Winch, Donald: 1978, *Adam Smith's Politics*, Cambridge University Press, Cambridge.

Research School of Social Sciences
 Australian National University
 GPO Box 4
 Canberra, ACT 2601
 Australia