1 Asymmetries

1. Aims

One of the most puzzling things about time is its flavor of asymmetry. We speak of time's "arrow" and "flow", trying to capture this feature—intending to suggest a profound difference between past and future. However, it isn't easy to do without these metaphors—to say literally what is meant by supposing that time has a direction in order to settle the question of whether or not it really does have one. Nor is it clear what we should make of pervasive temporal asymmetries like our capacity to control the future, the prevalence of decay, and the relative ease with which we can obtain knowledge of the past. Such asymmetries in time call for explanation, but do they indicate that time *itself* is asymmetric?

The atmosphere of mystery surrounding 'the direction of time' is crystallized in many specific philosophical and scientific problems. Let me mention three examples, and then give a more systematic account of the issues before us.

An especially intriguing problem concerns time travel. Is it possible to 'go back in time'? If so, how could it be done? And if the necessary technology will ever become available, why haven't we yet encountered visitors from the future? These questions are particularly tantalizing, as the topic of time travel is no longer confined to the realm of pure fantasy, having gained a measure of scientific respectability in the work of Kurt Gödel. It had been known since the acceptance of the Special Theory of Relativity that a form of time travel into the *future* could be accomplished by exploiting the fact that moving clocks run slowly. Thus someone could go on a very fast rocket trip and age, biologically, only two years, yet find on his return that ten years of Earth time had gone by. This sort of thing may or may not be properly called "time travel". In any case its possibility is uncontroversial. Much more problematic, however, is the idea of time travel into the past. Gödel's striking contribution was the discovery of certain spacetime structures, consistent with the General Theory of Relativity, that would allow journeys into history: "... by making a round trip on a rocket ship it is possible in these worlds to travel into any region of the past, present and future and back again, exactly as it is possible in other worlds to travel to distant parts of space" (1959, p. 560). But is this really conceivable? For how many of me would there be if I went back to 1950? Could I shake hands with myself? Could I change history—do something that in fact was not done? Worse still, what would stand in the way of 'autofanticide'—killing myself as an infant—an action whose failure would seem to be a necessary condition for success? We are left, therefore, with the following general question: Must we dismiss Gödel's results as mere mathematical curiosities, or can we dissolve the apparent paradoxes and preserve the possibility of time travel?

A more mundane problem involving time asymmetry concerns human motivation. What is the logic of rational action? What factors make one of a person's choices more reasonable than his other alternatives? This question interacts with the concept of time direction in that our processes of deliberation are *future* oriented. We try to figure out what we can do to maximize the chances of future benefits, and we don't much worry about the past, except perhaps for bouts of regret and self-congratulation. But why is this? Why is deliberation biased with respect to time? The explanation might seem obvious: namely, that causal influence works toward the future. We can never affect the past, so there's no point in planning for it. But there is another more subtle answer that cannot be dismissed. In the course of deciding what to do, we entertain various options. We find that our beliefs about what will transpire vary along with these different suppositions, but our conditional beliefs about what has happened remain constant. So, as far as the past is concerned, how can one action seem preferable to another? Thus our 'conditional belief asymmetry' suggests an alternative explanation of why past-oriented desires do not motivate us-one that does not invoke the direction of causation.

This pair of apparently conflicting answers to the question of why we act only for the sake of the future reflects two general conceptions of what makes something worth doing. On the one hand, there is the familiar, prevalent view—that the choiceworthiness of an act stems from what it might *cause*. On the other hand, and more liberally, we might take into account any event that the act would be *evidence* for. That is to say, even those events that are *not* caused by the act, but merely correlated with it, may be said to contribute towards its choiceworthiness.

Now one might wonder if there is any real dispute-or even any

real difference-between these conceptions. But the answer is yes, and Newcomb's notorious decision context clearly exposes their divergence. Imagine that a superior Being is able to foretell human choices, reliably and accurately, by means of a fancy psychological theory. Now suppose you are informed that vesterday the Being calculated whether or not you will now scratch your head, and that he set aside for you a valuable gift if and only if he predicted that you would scratch. Do you have good reason to do so? The causal conception of motivation says no. For it is not the act of scratching-but rather its precursors, recognized as such by the Being-that would cause the benefits. The evidential theory says yes. For given one's knowledge of the Being's abilities and intentions, scratching does raise the probability of benefits, even though it cannot bring them about. So the theories disagree. And pretheoretical intuitions don't settle the matter; they differ wildly from one person to another. Thus we are faced with three connected questions: Which is the right account of rational choice? Which action should be performed in Newcomb's dilemma? And which explanation of the decision time asymmetry is correct—why do we act for the sake of the future and not for the sake of the past?

These sample problems are philosophical insofar as their solutions involve clarification, examining relationships between concepts, and the unraveling of confusion. But not all our puzzlement about 'the direction of time' may be treated philosophically. Some of it reflects the need for a better *scientific* understanding of temporally asymmetric phenomena. We want to know why entropy tends to go up-why, in other words, highly ordered states decay but do not spontaneously evolve. This sort of asymmetry is commonly said to be, or to result from, the "arrow of time". But perhaps it has nothing at all to do with time. Perhaps time itself is perfectly symmetric, and increasing entropy is caused by cosmological conditions that do not bear on the nature of time itself. Moreover we are interested in what follows from the profusion of decay processes, and not just in the conditions that bring about this phenomenon. For example, it would be surprising if our impressive knowledge of the past (compared to our vast ignorance of the future) were unrelated to those physical time asymmetries. In addition there is the future-directedness of causal influence to account for. Many philosophers have thought that the scarcityperhaps impossibility-of backward causation is another deep effect of entropy growth.

The purpose of the following study is to investigate the entire cluster of questions—including those just mentioned—often lumped together under the heading "the problem of the direction of time."

4 Asymmetries

The project will involve (1) a clear specification of what it is for time to be asymmetric (directed, anisotropic), (2) a precise characterization of several pervasive temporally asymmetric phenomena, such as the fact that we are able to influence the future but not the past, and the fact that we know a great deal more about the past than the future, (3) an attempt to explain these asymmetries in a unified way and to examine their bearing upon the directedness of time itself.

In what immediately follows I shall describe ten apparent asymmetries that provide the main stimuli for our inquiry. Then I shall indicate the sort of understanding of these phenomena that will be our goal in subsequent chapters.

2. Ten temporally asymmetric phenomena

Now

We have a sense that time flows. We recognize a one-dimensional continuum of instants at which events are temporally located. But in addition there seems to be a kind of gliding index—now—that gradually moves along this array in the direction from past to future. "Time is the moving image of eternity," said Plato. "It is as if we were floating on a river, carried by the current past the manifold of events which is spread out timelessly on the bank." Or as Ovid put it, from a different point of view, "Time glides by with constant movement, not unlike a stream, for neither can a stream stay its course, nor can the fleeting hour."

This idea is captured less metaphorically in McTaggart's (1908) widely held theory about what would be required for the existence of time. According to McTaggart, the world contains a sequence of events ordered by such relations as *later than* and *simultaneous with*. But this would not suffice, he thought, for time to be real. In addition, to provide for genuine change in the universe, there must be a series of temporal specifications—distant future, near future, now, past, and so forth—with which events may be located, and which slides along the sequence of events in such a way that *now* applies to continually later and later events. McTaggart believed that, although essential to the existence of time, this 'motion of *now*' would be self-contradictory (since every event would have to possess the incompatible attributes of being past, present, and future). He concluded therefore that time is unreal.

This conclusion is literally incredible. Nevertheless, there is much to be learned from McTaggart's line of thought. We might well be favorably impressed with just the second component of his argument. We might agree (though this will take some showing) that the 'moving *now*' conception of time is indeed incoherent. But we might suppose this to mean, not that time is unreal, but that time need not (indeed *cannot*) be associated with a 'moving *now*'. If this is correct—if the 'moving *now*' is really just an illusion—then a further problem begins to loom. Why does this illusion have such a hold over us? Why is there such a strong sense that time does pass in a particular direction if it really doesn't?

Truth

Following Aristotle, it is often maintained that contingent statements about the future have no truth value, unlike statements concerning the past and present which are determinately either true or false. A prediction that war will break out next year will attain a truth value only then, when the event occurs or fails to occur. But right now there is no fact of the matter; for if there were, the presence or absence of the war would now be fixed, and nothing could be done to influence it. This contention is intended to imply a 'tree model' of reality whereby the past is petrified, uniquely fixed, over and done with, but the future contains a branching manifold of undetermined possibilities. And there is a tendency to invoke the 'moving now' in order to explain how a definite path up the tree is selected. It was Aristotle's belief that such an ontological distinction between the future and the past would have to exist if determinism and fatalism are to be avoided. We must assess this view and understand more precisely its relationship to the 'moving now' conception of time.

Laws

Turning now from metaphysics to science, suppose there is a process whose temporal mirror image is impossible-that is, a sequence of states of affairs, ABCD, where the reverse sequences, DCBA, would be ruled out by laws of nature. In such a case ABCD is said to be nomologically irreversible, and the operative laws are said to be timeasymmetric. Most of the theories that have ever been seriously entertained are not of this sort. They are like Newtonian mechanics in being time-symmetric and permitting the temporal reverse of all the processes in their domain. For example, imagine the motion of billiard balls colliding with one another on a frictionless table. If a movie of such a process were shown in reverse, no violation of law would be apparent. On the other hand, consider the second law (so-called) of thermodynamics, which states that the entropy of an isolated system will never decrease. This is evidently not time-symmetric, as there are processes of entropy increase (e.g., milk spreading through a cup of coffee) whose temporal inverses are ruled out by it. Although

thermodynamics has been superseded, the prospect of nomological irreversibility has not been banished from science. There is evidence that certain forms of fundamental particle decay (of the neutral K meson) are nomologically irreversible.

Our study of the asymmetry of time will involve an investigation of several questions provoked by the concept of irreversibility. Why, for example, is it so readily taken for granted that irreversible phenomena guarantee the asymmetry, or anisotropy, of time itself? This thesis will have to be justified in light of a clear statement of what it is for time to be asymmetric. Also, what is meant by the "temporal mirror image" of a process? This isn't at all obvious. One cannot simply say, as we have just blithely assumed, that the mirror image of ABCD is DCBA-the very same constituents in reverse order. For imagine a man eating dinner: soup, meat, desert, and then coffee. Must we hold that the reverse process is simply a meal in which coffee comes first? Then there is the matter of 'direction'. Would the presence of time asymmetry have any bearing on the so-called 'direction of time'? Why do we single out one of time's two directions for the title "the direction of time"? And is it conceivable that time should change direction?

De facto irreversibility

There are many processes whose temporal inverses are possible, although in actuality they never, or hardly ever, occur. Schematically the sequence of states *ABCD* is common, whereas instances of *DCBA* are extremely rare. For example, if a gas is concentrated in some small part of a container, it will expand to fill up the whole space available to it. But a gas that initially occupies the whole of its container never spontaneously shrinks into one corner. Similarly a source of light will emit a spherical beam that radiates outward; however, it never happens that a concave spherical beam converges toward a single point. In such cases we are dealing with *de facto* oneway processes whose inverses don't happen, though they are not precluded by the laws of nature.

Evidently some of the very problems arise here that I have just mentioned in connection with nomological irreversibility: problems that concern the meaning of "temporal inverse" and the conditions for the asymmetry of time itself. We shall examine arguments, on the one hand, that mere de facto irreversibility suffices to confer anisotropy upon time (Grünbaum 1963) and, on the other hand, that the association of irreversibility with the directionality of time is a mistaken dogma (Earman 1974). However, the central problem presented by these de facto one-way processes is to *explain* the temporal asymmetry that they display. It will be seen that laws of nature are not enough to account for it, especially given Boltzmann's reduction of thermodynamics to statistical mechanics. This means that the high frequency of entropy-increasing decay processes must be attributable to certain de facto conditions of the universe. Our aim will be to identify these conditions and to find their cosmological basis.

Knowledge

We know more about the past than we know about the future. For example, it is much easier to describe yesterday's weather than to forecast tomorrow's, and we know when the last five earthquakes in California took place, but not when the next five will be. Admittedly our capacity to predict in certain areas is quite impressive, and there are huge gaps in our knowledge of history; so it is hard to give a precise characterization of the asymmetry without exaggerating it. Nevertheless, it seems undeniable that there exist in the present many traces of earlier circumstances but relatively few reliable and recognizable indicators of what is to come. Thus we must acknowledge a dramatic difference in epistemological accessibility between the past and the future regions of time.

In trying to explain this time asymmetry, we shall look at various possible causes of it: (1) our capacity to act freely, and thereby to affect the future unpredictably; (2) our inability to go back in time and verify historical claims, which allows us to think mistakenly that our knowledge of the past is superior; (3) the direction of causation, which enables causal traces of the past but none of the future; (4) again, the direction of causation, because the meaning of the word "know" prohibits knowledge of any occurrences that do not cause our awareness of them; (5) the fact that, although the future may be physically determined by present events, the past is physically overdetermined, so current conditions provide several independent determinants of what has happened; and (6) the second law of thermodynamics, which allows us to infer, given the observation of a highly ordered system, that it has previously interacted with its environment. We shall find that none of these explanations is good enough. My analysis will proceed by examining the general nature of actual recording systems, such as books, photographs, and memory, that give us our special knowledge of the past, and then asking why it is that the temporal mirror images of such systems do not occur.

Causation

Effects seem never to precede their causes. We can influence the future but not the past. In other words, backward causation is impossible, de facto nonexistent, or, at the very least, extremely rare in this part of the world. The point of my weak characterization of the phenomenon is to avoid controversial presuppositions. For even if backward causation *sometimes* occurs, it remains to be explained why the *predominant* direction of causation is toward the future.

It is tempting to say: "A cause is, by definition, earlier than its effects." But this account must face up to several objections. Does it not trivialize the absurdity of acting for the sake of past events? For can we really suppose that the extreme irrationality of such a retroactive policy is merely a matter of stipulation? And what about cases of simultaneous causation? There appear actually to be such cases, yet they would straightforwardly falsify the alleged definition. Moreover is it right to consign backward causation to the same realm of inconceivability as the married bachelor or the weekend in which Sunday comes before Saturday? After all, we are able to imagine circumstances in which it might be tempting to say that an effect has occurred before its cause. To accommodate these difficulties, we shall have to take seriously certain alternative analyses of causation, accounts whereby its directional character does not stem purely from definition but depends on various contingent features of the world.

On the question of whether an effect can *ever* precede its cause, the issue turns largely on the merits of a well-known line of reasoning: namely that any backward causation hypothesis would automatically be refuted simply by waiting for an occasion on which the alleged effect is not present and then producing the alleged cause. I will try to show that this objection—sometimes called "the bilking argument" —is not as powerful as it is often taken to be. And I will apply this conclusion to three physical theories that postulate backward causation. Specifically I shall assess its effect on the plausibility of superluminal signals (tachyons), Feynman's identification of positrons with electrons moving backward in time, and Gödelian spacetime. There is plainly a close similarity between the bilking argument and some of the paradoxes of time travel that I mentioned at the start.

Explanation

We seek explanations of phenomena in terms of antecedent, rather than subsequent, circumstances. Indeed, explanations are sometimes roughly defined as accounts of an event that show how it could have been predicted—how, given the prevailing conditions, it was only to be expected. On the other hand, so-called teleological and functional explanations, which purport to explain a system's present state in terms of the attainment of some future goal, are of dubious scientific respectability. Why did the chicken cross the road? Presumably because of its earlier condition of wanting to reach the other side and not because of its later state of being there. Why is space threedimensional? Some physicists (adverting to what is known as the Anthropic Principle) say the reason is that, otherwise, stable planetary orbits would be impossible, and so life—and our awareness of the three-dimensionality of space—could not have evolved. But this sort of account, unsupplemented and taken at face value, can easily strike one as fundamentally misconceived, blatantly putting the cart before the horse.

It is plain, I think, that our ideas about the proper direction of explanation are intimately bound up with the directionality of cause and effect. What is not so clear, however, is precisely how these phenomena are related. A natural answer is that *explaining* is a matter of *specifying causes* and consequently that the explanation asymmetry is a product of causal directionality. But there is an interesting alternative to this approach. Reichenbach (1956) and Dummett (1964) have expressed the view that the arrow of explanation stems from the prevalence of de facto irreversible processes and that the direction of causation then derives from our concept of *cause* as *explainer*.

Counterfactual dependence

If the present were different from the way it is, then the future would be different. Thus true counterfactual conditional statements of the form 'If A had not occurred, then C would not have' seem to be about what would have happened subsequently if some actual event had not taken place, and not about what would have happened before.

For example, suppose the actual facts in some situation are (1) that Jones was on the roof of a high building, (2) that there was no safety net underneath him, (3) that he did not jump, and (4) that he was not hurt. A hypothetical negation of 3—the supposition that Jones jumped—would be taken to imply the negation of a subsequent fact, 4. In other words, we accept

If Jones had jumped, he would have been hurt

On the other hand, that same supposition is not easily taken to imply an alteration in preceding conditions. We do not normally conclude from the fact that Jones was on a high building and did not get hurt

If Jones had jumped, then there would have been a safety net underneath him

However, having noted the apparent time asymmetry of counterfactual implication, we must acknowledge that the issue is far from clear. One can quite well imagine someone saying

Jones would have jumped, only if there had been a safety net

And does this not mean

If Jones had jumped, there would have been a safety net

which implies

If Jones had jumped, he would not have been hurt

contradicting the initial claim?

In order to sort out this tangle of conflicting intuitions—and to determine whether such conditionals are really time-asymmetric —we must clarify the meaning of the counterfactual "If... then..." and specify its inferential relationships to other concepts. In particular, *causation* is a closely affiliated notion, and it will be important to understand precisely how it interacts with counterfactual dependence. The leading current approach to these matters is due to David Lewis (1973a, 1973b, 1979b), who argues with great flair that counterfactual dependence *is* time-asymmetric and that the direction of causation is an immediate consequence of this phenomenon. However, as we shall see, Lewis's approach is afflicted with a multitude of serious difficulties. It will be worth exploring the possibility that his theory inverts the actual explanatory relationship, and that counterfactual dependence grows out of the causal/ explanatory order, rather than vice versa.

Decision

We act for the sake of the future, not the past. More precisely, we would think it gravely irrational for someone to do something in order to ensure, or make probable, the occurrence of some desirable past event—or to preclude an undesirable one. It's no use crying over spilt milk, we say. In contrast, the future appears to be substantially controllable, and it is often reasonable, when some event of significance to us is in question, to be guided by whether any of our alternative actions would seem to make probable its occurrence.

One very natural explanation of this asymmetry is that we think it rational to act only for the sake of things we might *cause* to occur, and we are well aware that events, including our own actions, can exhibit causal influence only over the future. However, it is possible to resist this so-called *causal* decision theory. Instead, one might argue that the asymmetry results from the fact that our actions are probabilistically independent of prior circumstances. As I said at the outset, the conflict between these points of view comes to a head in the context of Newcomb's problem, and so this will be the focus of our discussion.

Value

We care a great deal more about what *will* happen to us than about what *has* happened. As a consequence we would much rather have pleasant prospects ahead and bad times behind us than the other way around. In particular, we dread death—future time at which we will not be alive—yet are quite unperturbed by the corresponding fact about birth—the reaches of past time when we were not alive.

This bias toward the future—the special importance we attach to what is still in store for us—has been noted by Derek Parfit (1984), who considers the question of what could account for it. I shall be concerned not only with that issue but also with the problem of how this bias is related to other temporally asymmetric phenomena. Is there, for example, any truth in George Schliesinger's (1980) thesis that the existence of this bias provides strong evidence in favor of the 'moving *now*' conception of time?

3. Explanatory maps

The general aim in the following chapters is to study each of these ten alleged asymmetries in order to discover: (1) which of them are genuine, (2) why they occur (3) what are the explanatory relationships, if any, between them, (4) whether there is some fundamental asymmetry (or two, perhaps) from which the others follow, and (5) what relation the asymmetries bear to the thesis that time itself is anisotropic (has intrinsic directional character). Are any of them constitutive of that thesis? Do they provide evidence for it? Do we need to invoke anisotropy in order to account for them?

To a large extent these objectives may be achieved by organizing the phenomena in a flowchart representing the explanatory relationships between them. As far as I know, the problem of the direction of time has not been approached before in this way and on such a broad front, and there is no explicit proposal of the sort of explanatory map I have in mind. Nevertheless, it is possible to extract from the writings of those who have struggled with these issues some partial theories. As an illustration I shall briefly describe and comment on the partial map (figure 1) implicit in Reichenbach's (1956) work.

According to Reichenbach, de facto irreversible processes consist in the creation and gradual decay of order. The observation of a highly ordered (low entropy) state tells us (given the second law of therKnowledge

Figure 1

modynamics) that since the chances are negligible of spontaneous evolution into such a state, the system must have interacted in some characteristic way with its environment. Thus we obtain information about the past. In addition, by reference to the maximally ordered state, we may confer high probability upon the subsequent states of partial order—states that would otherwise seem amazingly coincidental and improbable—and thereby explain them. Thus the direction of explanation is tied to the direction in which order is dispersed. Moreover, assuming that our notion of *cause* is *that which explains*, we also derive the direction of causation indirectly from the orientation of irreversible processes.

This line of thought is extremely ingenious, yet controversial in every detail. For example, Mackie (1974) maintains that we should get the existence of irreversible processes from the direction of causation, rather than the other way around; Earman (1974) suggests that the knowledge asymmetry stems from the causal asymmetry, and that neither has much to do with entropy; von Wright (1971) says that the temporal orientation of causation comes out of our ability to manipulate the future; Salmon (1984) argues that explanation should be defined as a specification of causes.

Thus there is strikingly little agreement about the sources of temporally asymmetric phenomena and about the interdependencies among them. This is in some part because philosophers have tended to approach these questions in an overly piecemeal way. Consequently their conclusions are often undermined by a failure to appreciate and accommodate the needs of a comprehensive account. The following investigation will attempt to avoid this shortcoming. It will offer explanations of the asymmetries and criticism of alternative proposals. And its theses will gain credibility from their interaction within an unusually broad conceptual network.

The overall theory towards which I shall be working involves the idea that time *itself* has no intrinsic directionality or asymmetry, and it explains the temporally asymmetric phenomena accordingly, as shown in figure 2. To begin with, the de facto irreversible processes are given a cosmological explanation (in terms of the randomness of microscopic conditions following the big bang) and are employed to account for the fact that we know so much more about the past than

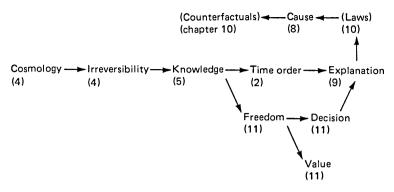


Figure 2

about the future. This knowledge asymmetry, by virtue of the difference between memory and expectation, yields our conception of succession. Time order is built into the notion of explanation in such a way as to imply that earlier facts may be taken to explain later ones. And the direction of physical explanation yields the direction of causation, since explanation is a description of causes. Counterfactual dependence is then analyzed in terms of causal explanation and, contrary to first appearances, it turns out not to be time-asymmetric after all.

Moreover an element in our sense of free choice is that the knowledge of what we are going to do evolves through a process of deliberation, intention, and action. Given the knowledge asymmetry, such a process must have a particular temporal orientation: namely, deliberation, followed by intention, followed by action. This ordering implies that our beliefs about what will occur in the future are sensitive to variations in what we suppose we will do. On the other hand, since the relationship between past events and our prospective action is mediated by the beliefs and desires that we recognize during deliberation, there is nothing we can normally infer about the past from some supposed action that we cannot already infer during deliberation independently of any such supposition. This is what gives rise to the decision asymmetry-our tendency to act for the sake of the future-which in turn plays a role in fixing the direction of explanation. The value asymmetry-the special importance we attach to future experiences—is also explained by the temporal orientation of our sense of freedom. Given that the typical decision process involves desire, deliberation, decision, action, and fulfillment (in that temporal order), then a desire for future satisfaction will be an aid to survival. For this attitude increases the chances that future selfish desires will be fulfilled; whereas there is no mechanism by which a present wish that past desires have been satisfied would be associated with any increased fulfillment of those past desires.

Needless to say, any sketch such as this is likely to be dangerously oversimplified. In particular, one should guard against supposing that the arrows in figure 2 always indicate the same type of explanatory connection. However, I hope, at least, that it is helpful to see at a glance the rough shape of the theory that will gradually emerge. A more detailed summary is given at the end of the book.

It is natural to think that time is obviously asymmetric, with an indefinable, yet undeniable, directional character, and to think that this is responsible in one way or another for the various asymmetries that we have just been discussing. But as I have indicated, my view is very different. The first step in presenting it will be to attack the idea that time has a direction (chapter 2), or indeed that it is asymmetric at all (chapter 3). Then, after discussing the behavior of entropy, I shall turn to an affiliated phenomenon, called "the fork asymmetry", which is closely related to the fact that regularly associated events must have a common cause but need have no joint effect. This phenomenon, I will argue, does provide a source for many of the asymmetries we are dealing with. The nature and limits of this dependence, and the role of other factors, are elaborated in subsequent chapters.