

I. CHOICE AND INDETERMINISM

Weigh(t)ing Reasons

Making some choices feels like this. There are various reasons for and against doing each of the alternative actions or courses of action one is considering, and it seems and feels as if one could do any one of them. In considering the reasons, mulling them over, one arrives at a view of which reasons are more important, which ones have more weight. One decides which reasons to act on; or one may decide to act on none of them but to seek instead a new alternative since none previously considered was satisfactory.¹

After the choice, however, others will say we were caused to act by the considerations which were (or turned out to be) more weighty. And it is not just others. We too, in looking back at our past actions, will see which reasons swayed us and will view (accepting) those considerations as having caused us to act as we did. Had we done the other act, though, acting on the opposing considerations, we (along with the others) would have described those considerations as causing us to do that other act. Whichever act we do, the (different) background considerations exist which can be raised to causal status. Which considerations will be so raised depends upon which act we do. Does the act merely show which of the considerations was the weightier cause, or does the decision make one of them weightier?

The reasons do not come with previously given precisely specified weights; the decision process is not one of discovering such precise weights but of assigning them. The process not only weighs reasons, it (also) weights them.² At least, so it sometimes feels. This process of weighting may focus narrowly, or involve considering or deciding what sort of person one wishes to be, what sort of life one wishes to lead.

What picture of choice emerges if we take seriously the feeling that the (precise) weights to be assigned to reasons is "up to us"? It is causally undetermined (by prior factors) which of the acts we will decide to do. It may be causally determined that certain reasons are

reasons (in the one direction or the other), but there is no prior causal determination of the precise weight each reason will have in competition with others. Thus, we need not hold that every possible reason is available to every person at every time or historical period. Historians and anthropologists delineate how certain ideas and considerations can be outside the purview of some societies, some of whose reasons would not count as reasons for us. (Yet, there does remain the question of whether an innovator couldn't have recognized as a reason something outside the purview of others in his society.) Psychology, sociobiology, and the various social sciences, on this view, will offer casual explanations of why something is or is not a reason for a person (in a situation). They will not always be able to explain why the reasons get the precise weights they do. Compare the way art historians treat style; not every style is equally available to every artist in every period, yet within a style creative choices are made, and some artistic revolutions introduce new stylistic possibilities.

It is neither necessary nor appropriate, on this view, to say the person's action is uncaused. As the person is deciding, mulling over reasons R_A which are reasons for doing act A and over R_B which are reasons for doing act B, it is undetermined which act he will do. In that very situation, he could do A and he could do B. He decides, let us suppose, to do act A. It then will be true that he was caused to do act A by (accepting) R_A . However, had he decided to do act B, it then would have been R_B that caused him to do B. Whichever he decides upon, A or B, there will be a cause of his doing it, namely R_A or R_B . His action is not (causally) determined, for in that very situation he could have decided differently; if the history of the world had been replayed up until that point, it could have continued with a different action. With regard to his action the person has what has been termed contra-causal freedom—we might better term it contra-deterministic.*

* The notion of contra-causal human freedom (though not the term) originated with Philo. In his view, God, in creating the world, reserved for himself the power to upset laws by working miracles, and gave to man a portion of that same power—although man's 'miracles' are not worked with respect to laws that he himself created. (See Harry A. Wolfson, *Philo*, Harvard University Press, Cambridge, 1947, Vol. I, pp. 431, 436.) The Epicureans denied causality altogether, and Chrysippus held that causality, by its nature, stopped at the will of man. The Philonic view is the first to place absolute

Thus, we draw a distinction between an action's being caused, and its being causally determined. Some philosophers would deny this distinction, maintaining that whenever one event causes another, there holds a general law in accordance with which it does so: some specification of the first event (along with other conditions which hold) always is and would be followed by an event of the same type as the second. It is a metaphysical thesis that the root notion of causality, producing or making something happen, can operate only through such lawlike universality. If this were correct, and if a law could not hold only at that (moment of) time, then causality necessarily would involve causal determination: under exactly the same conditions repeated, exactly the same thing would have (again) to happen. According to the view that distinguishes causality from causal determination, an act can be done because of something and have a cause even though in exactly the same conditions another act could have been done. It is common, in retrospect, to see what caused us to act as we did. Although we can retrospectively identify a cause, this does not mean our action was causally determined; had we acted differently in that situation (as we could have) we retrospectively would have identified a different cause— R_B instead of R_A .

The weights of reasons are inchoate until the decision. The deci-

free will within a world of some causality which otherwise would apply and which is suspended. (See H. A. Wolfson, *Religious Philosophy*, Harvard University Press, Cambridge, 1961, p. 196; *The Philosophy of the Kalam*, Harvard University Press, Cambridge, 1976, p. 733.) It became a matter of controversy within Christian theology whether humans naturally retained this gift of free will from God, or whether, after Adam's fall, God withdrew it as a matter of course and bestowed it only as a matter of divine grace. (See H. A. Wolfson, "St. Augustine and the Pelagian Controversy" in his *Religious Philosophy*, pp. 158–176.)

It is instructive to apply to these matters the notion (developed in Chapter 2 above) of an egalitarian theory, wherein there is marked out a natural state, deviations from which have to be explained by special forces or reasons. There are at least three views: (1) Philo's view: man has free will as a gift from God, a gift bestowed in creating man's nature; hence free will is man's natural state, and could not be altered by Adam's sin; (2) man's natural state was unfree, but in one act God gave all men free will as a donation of some of his powers; (3) in response to Adam's act God altered man's natural state (if 1 had been true) or revoked his general gift; he now has to bestow free will upon each person individually. (This last is Augustine's view.) Notice that proponents of each of these views can agree that all people have free will, yet disagree about its explanatory status or explanation.

sion need not bestow exact quantities, though, only make some reasons come to outweigh others. A decision establishes inequalities in weight, even if not precise weights.³

These bestowed weights (or comparative weightings of reasons) are not so evanescent as to disappear immediately after the very decision that bestows them. They set up a framework within which we make future decisions, not eternal but one we tentatively are committed to. The process of decision fixes the weights reasons are to have. The situation resembles that of precedents within a legal system; an earlier decision is not simply ignored though it may be overturned for reason, the decision represents a tentative commitment to make future decisions in accordance with the weights it establishes, and so on.⁴

The claim that we always do what we most prefer or always act from the strongest motive is sometimes said to be empty of content, since the preference or the strength of motive is identified by what the person does. If the claim is to have empirical content, it must sometimes be possible to discover what a person's preference or strongest motive is via some other situation, to independently identify it in order then to check in this situation whether the person is doing what he most prefers or has the strongest motive to do.⁵ Defenders of the claim do point out other situations (of choice or answering questions) where the relevant preference or motive can be identified; so the truth of the claim in this decision situation is testable, given the assumption that the preference or motive is stable from the one situation to the other.⁶ However, if our conception of the bestowal of weights (with a commitment that lingers) holds true, then these independent "tests" are to be interpreted differently. We do not always act on what was a preexistingly strongest preference or motive; it can become strongest in the process of making the decision, thereafter having greater weight (in other future decisions) than the reasons it vanquished. The prior independent test of a preference therefore need not discover one that existed; it may establish a preference which then consistently carries over into a new decision situation. The testing procedure cannot show that we always act on a preexistingly strongest preference or motive.⁷

Only when there are opposed reasons for different actions is it necessary to arrive at a weighting; otherwise, one can just do what all the reasons favor. However, neither group of these opposed reasons

need be moral; decisions that involve a conflict of duty or other moral motives with (nonmoral) desires are only a subclass of the free decisions.⁸ Shall we say, though, that every free decision involves a conflict of some sort, with reasons pulling in different directions? The reasons in conflict need not then have indeterminate weight, for a free decision may "act out" an earlier weighting decision as precedent. (But is there always present a reason of indeterminate weight to reexamine and overturn an earlier precedent, which reason itself must be given a determinate lesser weight in the decision to follow the precedent?) Even though it will include no interesting cases we especially want to judge, still, we may formulate the theory to avoid the uncomfortable consequence that actions in the face of no contrary reasons are not free ones.

Is this conception of decision as bestowing weights coherent? It may help to compare it to the currently orthodox interpretation of quantum mechanics. The purpose of this comparison is not to derive free will from quantum mechanics or to use physical theory to prove free will exists, or even to say that nondeterminism at the quantum level leaves room for free will. Rather, we wish to see whether quantum theory provides an analogue, whether it presents structural possibilities which if instanced at the macro-level of action—this is not implied by micro-quantum theory—would fit the situation we have described. According to the currently orthodox quantum mechanical theory of measurement, as specified by John von Neumann, a quantum mechanical system is in a superposition of states, a probability mixture of states, which changes continuously in accordance with the quantum mechanical equations of motion, and which changes discontinuously via a measurement or observation. Such a measurement "collapses the wave packet", reducing the superposition to a particular state; which state the superposition will reduce to is not predictable.⁹ Analogously, a person before decision has reasons without fixed weights; he is in a superposition of (precise) weights, perhaps within certain limits, or a mixed state (which need not be a superposition with fixed probabilities). The process of decision reduces the superposition to one state (or to a set of states corresponding to a comparative ranking of reasons), but it is not predictable or determined to which state of the weights the decision (analogous to a measurement) will reduce the superposition. (Let us leave aside von Neumann's subtle analysis, in Chapter 6, of how any placing of the

"cut" between observer and observed is consistent with his account.) Our point is not to endorse the orthodox account as a correct account of quantum mechanics, only to draw upon its theoretical structure to show our conception of decision is a coherent one. Decision fixes the weights of reasons; it reduces the previously obtaining mixed state or superposition. However, it does not do so at random.

Nonrandom Weighting

Granting the coherence of the conception wherein the process of decision bestows weights, still, is that free will? An action's being non-determined, we saw earlier, is not sufficient for it to be free—it might just be a random act. If we acted in the way uranium 238 emits alpha particles, determinism would be false but (unless we are greatly mistaken about uranium 238) we would not thereby have free will. What makes the bestowal of weights on reasons any different? If that too is a random act, then is acting on those weights in that very decision other than random? Acting on those same weights later will not be random, but is it better than any other determined act if it traces its history back not to causes before birth but to a recent random weighting of reasons?

How can the giving of weights be other than random? Since (by hypothesis) there is no cause for giving or bestowing these particular weights on reasons rather than other weights, must it be merely a random act when these are bestowed? (Let us leave aside for the moment the distinction between 'caused' and 'causally determined'.) If the absence of causation entailed randomness, then the denial of (contra-causal) free will would follow immediately. However, 'uncaused' does not entail 'random'. To be sure, the theorist of free will still has to explain wherein the act not causally determined is nonrandom, but at least there is room for this task.

In what way is the bestowal of weights not simply random? There may be causes limiting the reasons on which (nonzero) weight can be bestowed, and the interval within which these weights fall may similarly be limited. However, although it is not a random matter that the weights bestowed fall within this range, neither is that decided by the person. The question remains: how is her decision

among the alternatives causally open to her (the alternatives it is not causally determined she won't choose) not simply a random matter?

First, the decision may be self-subsuming; the weights it bestows may fix general principles that mandate not only the relevant act but also the bestowing of those (or similar) weights. The bestowal of weights yields both the action and (as a subsumption, not a repetition) that very bestowal. For example, consider the policy of choosing so as to track bestness: if the act weren't best you wouldn't do it, while if it were best you would. The decision to follow this policy may itself be an instance of it, subsumed under it.

Another issue shows how an act of decision can refer to itself. In contrast to optimizing models of decision, which see the agent as maximizing some objective function, Herbert Simon has presented a *satisficing* model of decision (to use his term): an agent will do an action that is "good enough", but failing to find one among his alternatives he will search for still others; repeated failure to find a suitable one will change his view of what is good enough, lowering his level of aspiration. It is natural to try to embed these considerations within an optimizing model that includes the costs of searching for new alternatives, gathering further information, as well as estimates of the probability of finding a new better alternative. The optimizing model would view 'searching for another alternative', or 'searching for more information about the other alternatives' as (always) among the actions or options already available. It therefore sees the choice among these available alternatives as involving maximization (under risk or uncertainty).¹⁰ This faces the following difficulty, however. In making that choice among those alternatives on the basis of that information, was the structuring of that choice situation based on a previous optimizing decision or upon a satisficing decision that the structuring was "good enough"? Whichever, is not a decision made, at some point, which includes estimates of the costs and benefits of gathering more information in *that very* choice situation? Won't there be some decision, whether optimizing or satisficing, whose scope covers all costs including its own?¹¹

Consider a self-subsuming decision that bestows weights to reasons on the basis of a then chosen conception of oneself and one's appropriate life, a conception that includes bestowing those weights and choosing that conception (where the weights also yield choosing that self-conception). Such a self-subsuming decision will not be a

random brute fact; it will be explained as an instance of the very conception and weights chosen. (I do not say that all of one's choices or all that bestow weights are self-subsuming in this way; however, the other ones that are based on weights previously given in such decisions, revokable weights, will inherit autonomy.) It will no more be a random brute fact than is the holding of a fundamental deep explanatory law that subsumes and thereby explains itself. (See Chapter 2 above.) A self-subsuming decision does not happen inexplicably, it is not random in the sense of being connected to no weighted reasons (including the self-subsuming ones then chosen). But although it doesn't happen just randomly, still, there are different and conflicting self-subsuming decisions that could be made; just as there are different fundamental, self-subsuming laws that could hold true, could have held true. Is it not arbitrary then that one self-subsuming decision is made rather than another? Won't it be left inexplicable why this one was made (rather than another one)?

Understanding and Explaining Free Choices

First, a word about explanation and intelligibility. When deductive explanations subsume an event under a covering law, then we understand why that event occurred rather than any other. (It is another question why that covering law held rather than another; given that it did, we understand why the particular event occurred.) The situation is different with statistical explanations. Suppose a fundamental law states that the probability that anything has property Q given that it has property P is .95; if we wish to explain why some entity has property Q, we cannot deduce this fact from the entity's having property P plus the probabilistic law. Nevertheless, many have thought the statistical law does enable us to explain why the entity has property Q. Hempel has held that high probability events are explained by subsuming them under probabilistic laws; the high-probability probabilistic explanation is an approximation to a deduction.¹² What of the low probability event, though; when we encounter an entity that is P but isn't Q, can we explain why it is not Q? There is no way to do this on Hempel's view. True, if many P's are observed, then it can be very likely that one or another of them will not have property Q; we expect (as our best estimate) only 95 per-

cent of the P's to have property Q, so there is a high probability that one of the very many P's we encounter will lack Q. When this one lacks Q, isn't the explanation simply that some (small) percentage of the P's will lack Q, and this is simply one of the ones that do? Strictly, on Hempel's view we have only an explanation that some P or other will lack Q, for that fact has a high probability; but we do not have an explanation of why, for example, this entity E lacks Q even though it has P. But even though we cannot deduce or yield with a high probability that it, entity E, will lack property Q, still, when we encounter it don't we know the explanation of why it does? We know there exists a system, a chance mechanism or whatever, that generates some P's that are not Q's, and we explain why this P is non-Q, by its being one of the things spewed forth by the operation of the chance mechanism.¹³ The alternative (if there are some fundamental probabilistic laws) is to say these low probability events are unexplainable.

The moral I wish to draw is this: we can have an explanation and understanding of why something occurred even when we do not know of any reason why it, rather than something else, occurred that time, in that instance. Even when the event is random, its occurrence need not be inexplicable; it can be seen as an event, one of a type to be expected, arising from a mechanism or system that, in a way we may have better or worse understanding of, yields such events among others.

I am not suggesting that free decisions are random happenings from a chance mechanism with a well-defined probability distribution (whether flat or otherwise) over the various alternative actions. The process of choice among alternative actions is different; there are not fixed factual probabilities for each action, there is no such dispositional propensity or limit of long-run frequencies or whatever. Rather, there is a process operating wherein each alternative action could be yielded, and one of them was. This time, the process gave rise to that particular alternative. (Compare: this time the random system yielded that particular event.)

To be sure, we do not want to say simply that there is a process which could give rise to any of the alternative actions—we want to know more about the process, we want to delineate and understand it, we want to know how it works.

According to the view currently fashionable, we adequately under-

stand a psychological process only if we can simulate that process on a digital computer. To understand a psychological notion is to know a set of quadruples that would place a Turing machine under the notion. Any process of choosing an action that could be understood in this sense would appear not to be a process of free choice. Suppose that this is so.¹⁴ Does the fact that we cannot, in this sense, understand what a free choice is, indicate some defect in the notion of a free choice or rather is the defect in the view that this mode of understanding is the sole mode? Is the result, that we cannot understand what a free choice is, an *artifact* of this method of understanding?¹⁵

In what other way, if not simulation by a Turing machine, can we understand the process of making free choices? By making them, perhaps. We might interpret those theorists who pointed to our choices not as trying to prove that we made free choices but as ostensibly explaining the notion, showing its intelligibility. Were they saying that we understand free choice and agency by virtue of making free choices as agents? To accept a (restricted) form of knowledge by self-acquaintance, encompassing knowledge of a mode of action and of ourselves, runs afoul of views that we know something only when (and to the extent that) we know the laws it obeys.¹⁶ However, even if such views are rejected, the nature of this other mode of knowledge, by self-acquaintance, is unclear;* an adequate theory, showing how it is possible, would take us into issues far removed from our present concern without helping us especially with the topic of free will. Our problem is that we are puzzled about the nature of free choices, so any inside knowledge we may have of such choices due to and in making them obviously hasn't served to clear up our puzzles about their nature. It is tempting to say our puzzlement stems from supposing we must be able discursively to say or describe what a free choice is like, yet the fact that we cannot, when we are directly acquainted with them, doesn't interfere with understanding them. But too many ineffabilities spoil the philosophical

* "How can we know in that way, without reflective conceptual scrutiny? And will not all the knowledge be in the reflective scrutiny?" For a presentation of a view that avoids this philosophical picture but leaves much obscure (at least as judged by the mode of knowledge it claims is not the only one) see Aurobindo, *The Life Divine*, Book II, pt. I, ch. X, "Knowledge by Identity and Separative Knowledge".

broth. Since I do not myself have even the feeling of understanding, I will continue the (discursive) attempts at explanation.

We can explain an action as an intentional doing arising out of a process of choice among alternatives, if we can illuminate this process; however, we need not offer a Turing machine model, a computer simulation of the process of decision that matches which alternative the person chooses. We have said already that the decision process (sometimes) bestows weights on the reasons for and against the various alternatives, and that this bestowal of weights is self-subsuming and so to that extent not random. Still, there can be different self-subsuming bestowals of weight. Although after one occurs we will be able retrospectively to give a reason as the cause (though without causal determination), can anything be said about why that one self-subsuming decision is made rather than another? No, the weights are bestowed in virtue of weights that come into effect in the very act of bestowal. This is the translation into this context of the notion of reflexivity: the phenomenon, such as reference or a law's holding, has an "inside" character when it holds or occurs in virtue of a feature bestowed by its holding or occurring.

The free decision is reflexive; it holds in virtue of weights bestowed by its holding. An explanation of why the act was chosen will have to refer to its being chosen. However, not every act you do is a minor miracle of reflexive self-subsumption, only the ones involving choice of fundamental principles and self-conception. (Yet since such a choice is revokable, do later choices reaffirm it, and so also involve reflexive self-subsumption?)

Suppose a process of decision can have these features, bestowing weights in a self-subsuming fashion which is reflexive. The decision then does not simply dangle there at random—we can see the many ties and connections it has (including internal ones); the particular decision is not inexplicable—we see it as something that could arise from a process of this sort.

More might be demanded, however; it might be demanded that the theorist of free will show how the decision is causally determined. Otherwise, it will be said, the character and nature of the decision will remain mysterious. But clearing up any mystery in that way would come at the cost of the act's contra-causal freedom. No adequate condition on explanation or understanding necessitates either causal explanation or Turing machine delineation. Free will is

to be explained differently, by delineating a decision process that can give rise to various acts in a nonrandom nonarbitrary way; whichever it gives rise to—and it could give rise to any one of several—will happen nonarbitrarily. These remarks are independent of the particular process we have delineated here, involving the bestowal of weights, reflexive self-subsumption, and so on. What is inappropriate is to demand that a free choice be explained in a way that shows it is unfree.

The theme of the bestowal of weights to reasons, in a situation of no preexistingly determinate weights, seems to me phenomenologically accurate¹⁷ and proper to emphasize. I have more worries about terming this bestowal nonarbitrary and nonrandom because it is self-subsuming and reflexive. This position has too much the flavor of applying shiny new tools and ideas everywhere, as a magic key—except that some of the applications depend, perhaps, upon these ideas being not so well understood, not so shiny. So we should be somewhat wary of this use of the themes of self-subsumption and reflexivity to delineate the nonarbitrary nature of a free choice. They do have the right flavor, though. For example, consider all the talk (in the literature) of "stepping back" to reconsider any previous commitment or self-conception. Is this merely the analogue of Peirce's point in epistemology that anything can be doubted but not everything at once—any motive or reason can be examined though not every one simultaneously? To where do we step back? In the case of a free choice, it seems appropriate that it be to somewhere such that (the act of) stepping to there is an instance of being there, which you are in virtue of a feature of your being there. "Stepping back", at least sometimes, is not like moving up to different levels in a type theory hierarchy; rather, it is self-subsuming and reflexive.

There are other issues that need to be explored, but will not be here: how the later (possible) revocation of bestowed weights works; whether there is causal leeway not only in bestowing weights on reasons, but also in the generation of alternative actions;¹⁸ how the later less fundamental choices, which spin out the previously bestowed weightings, inherit autonomy. One further word can be said about the commitment involved in the bestowal of weights. Acting later on those weights anchors your later choices to them, and them to the later choices. Part of this nonrandom character of the weighting is shown by the life built upon them; perhaps it not merely is

exhibited there but exists there.¹⁹ If this is too strong, at least we may see the later adherence to weights as an indication of their nonrandom character; if the choice of these weights was simply random and arbitrary, would they win continued adherence?

Could One Have Bestowed Otherwise?

In the first chapter we saw how, within a closest-continuer framework, the self weights dimensions to yield a measure of closeness for itself, in accordance with its own self-conception, and saw how this weighting, including of plans, desires, and values, can be an important component in reflexive self-synthesis. Another way in which bestowal of weight upon reasons can be nonarbitrary is that the self can synthesize itself around this bestowing: "I value things in this way." If in that reflexive self-reference, the I synthesizes itself (in part) around the act of bestowing weight on reasons, then it will not be arbitrary or random that *that* self bestowed those weights.*

The process of decision can yield the intentional doing of different actions, and it would have if different weights had been assigned, which could have happened. But does it follow that the person could have done otherwise, that it was within the person's power to bestow different weights, as opposed to that merely happening? In what way could the person have *done* otherwise, not merely been the arena in which otherwise happened?

It would be fruitless to embark upon the theoretical regress wherein a different intentional action of bestowing weights occurs with its own separate weights which have to be bestowed by a still separate act.²⁰ And why is it asked only if another bestowal could

* Does this have the consequence that that self could not have bestowed weights differently? Not if the bestowal itself receives only a limited weight in the actual self-synthesis, and if the alternative syntheses involving different bestowals would then have been the closest continuers of the same earlier self, the one the actual synthesized self most closely continued. By most closely continuing the same earlier self, the other (possible) syntheses around other bestowals would have been that self later, just as the actual synthesized self is. So the particular bestowal isn't essential to the self with the consequence that the self couldn't have done otherwise, yet as a weighted component of a self-synthesis, neither is it random and arbitrary in relation to that self.

have been done; why is it not similarly asked whether the bestowal that did occur was a doing or merely a happening? Maybe it is possible for weights somehow to just happen to get bestowed on reasons; however, when the bestowal is anchored and tied in the way we have described, to a formed self-conception (even if formed just then), if it is self-subsuming and reflexive, leading to later (revokable) commitment, then it is a doing, not a happening merely. If all that context and stage setting (compare Wittgenstein) does not make it an action, what alternative conception of action is being presupposed? The actual bestowal of weights on reasons is a doing and not merely a happening; another and alternative bestowal of weights on reasons could have occurred instead—this one wasn't causally determined, and others aren't causally excluded—with all of the accompanying context and stage setting appropriate to it, so that alternative bestowal too would have been a doing and not merely a happening. The person could have bestowed differently.

Why Free Will, and How?

We have sketched a view of how free will is possible, of how without causal determination of action a person could have acted differently (in precisely that situation) yet nevertheless does not act at random or arbitrarily. (I admit the picture is somewhat cloudy.) Are there considerations that make this view plausible, not proving it true but indicating enough plausibility so that we do have an admissible (possible) explanation of how free will is possible? We might try to place free will within an evolutionary framework, thereby making it scientifically respectable. This could occur even without understanding how free will works, how weight is bestowed on reasons; if we could understand its adaptive value, understand why once it came to exist it would be selected for, then it would be placed within the network of scientific theory.²¹

What is the adaptive advantage to bestowing weights upon reasons in a self-subsuming reflexive fashion with a result that is not determined causally, so that if this capacity arose (at random) and was (to some significant degree) heritable, it would be selected for in the evolutionary process? It will be plausible that we do bestow weights

on reasons (the particular bestowal not being causally determined) if both we naturally think (or feel) we do, and this would have been selected for, had it arisen, in the evolutionary process. It is not difficult to see the advantages for an organism of being able to behave in new ways in new situations, not just random thrashing and done only when it is appropriate or needed. The self-subsuming character of the decision may limit the thrashing; only some candidates (but more than one) will pass this test. But among these candidates, why is the choice not simply random, with an equal probability of choosing any one; what adaptive advantage does a nonrandom choice have over a random one here? Unfortunately I have not myself been able thus far to demarcate the special adaptive advantage free will might have; nonetheless, the evolutionary framework is a useful one to mark for further pursuit.

It is quite extraordinary that (some of) our decisions should escape the lattice of causal determination. Supposing or granting that there would be a function to their doing so, how does it manage to happen? What is it about (some of) our decisions, what feature do they have, that lifts them from the nexus of causal determination? Here, I can only offer a vague speculation. Reflexive self-subsuming acts have an intrinsic depth; the way they turn back on themselves, refer to themselves, refer reflexively "from the inside", gives them an unlimited (infinite?) semantic depth. Might it be that this cannot be caused, so the very feature that makes the decision nonrandom, reflexive self-subsumption, by its character eludes being caused by something else?²²

Why cannot something reflexively self-referring be causally determined; even if it has unlimited (semantic) depth, why can it not be caused by something else which is (at least) equally deep? Descartes utilized a principle—one I don't understand well—that an effect cannot have a greater degree of "reality" than its cause.²³ We seem to be dealing with a similar type of principle here, that an effect cannot have a greater semantic depth than what causally determines it; that aspect of the effect, its semantic depth, cannot be causally determined by anything of lesser semantic depth. Semantic depth doesn't arise out of nothing; at least, it is not causally determined out of nothing or even less than itself. This principle is more modest (and plausible) than one that says reflexively self-subsumptive semantic

depth cannot be causally determined at all, even by something of equivalent depth. Various philosophical controversies about reduction seem to hinge on this general principle; for example, whether intentional contexts can be reduced to nonintentional ones.²⁴ The principle, if correct, would delimit a barrier to explanatory or causal reductions: no reduction of the deep to the shallow.

Free decisions, then, would be an instance of this more general principle, a principle for whose own explanation we would have to seek deeper yet. More accurately, the principle says reflexive self-subsumptive decisions will not have causes of a certain sort, of lesser depth. But cannot one of mine be caused by someone's saying, "make that particular reflexive self-subsumptive decision or I will shoot you"?²⁵ Would this reference by the threatener to my decision have at least the same depth as the decision and so be able to cause it? Here, perhaps we have a further explanation of why theorists of coercion focus upon the interpersonal situation where someone else intentionally threatens you, placing their intentions about your action (and choice?) within your decision. In contrast to a fact of nature such as an electrical storm that gives the same probability distribution over relevant consequences, and to another's act that accidentally alters the probability distribution you face, this intentional interpersonal case has the requisite degree of depth to causally determine your (even semantically deep) act, perhaps. Still, one can decide to resist a threat, one can weigh it and go against it. Is one's decision deeper than the threat in that one "steps back" from it and weighs it? Yet why cannot the threat also refer to itself ("do that action on this very threat or else"), and so achieve the same depth as the decision that weighs it?²⁶ But can it achieve the same depth as a decision that weighs it?

The principle of nonreduction to lesser semantic depth, and no causal determination by it, is an interesting one, whose consequences (and possible explanations) are worth exploring. However, it is not a principle I can put forth with any confidence or accept yet, even very tentatively.²⁷ I mention it merely to help explain how a free decision might be possible, how it is possible that it is possible. We have not produced any very good reasons for thinking the principle true,²⁸ nor any for thinking that (some) indeterminism does hold true, as opposed to speculating how it might possibly hold.

Is Free Will Valuable?

How important is it to this picture that our decisions not be causally determined? Suppose that in certain types of situations, we did reconsider our weighting of reasons, our self-conception, and our lives, but the new position we arrived at was causally determined—we always would arrive at precisely that position in precisely those circumstances. How different would this really be from the indeterministic situation described earlier, especially if others (and we ourselves) could not reliably predict the new position? How significant is the difference between this deterministic situation and its indeterministic mate?

How does determinism threaten (to undercut) our value? One can have various pictures: our being simply an arena in which causes play out their effects,* or our being puppets—marionettes moved by the causes at the other ends of the strings. It will be pointed out that we are not extremely simple input-output devices, much internal processing takes place, involving feedback loops and other delightful “software”; however, does that not make us merely more complicated puppets, but puppets nonetheless? True, much of these causes occur “inside” us—is it better to be a hand puppet than a marionette?

If we are caused to be aligned with correct value, to track some value trait such as bestness or rightness, there will be a value we have in virtue of this. Being aligned with correct values, tracking them, itself is a valuable thing, the connection itself is valuable. (In the next chapter, we will want a theory of value to explain or have the consequence that this connection with value itself is valuable.) However, it seems this value is not due to us, even if we play some part in it. On the causal determination view, though the connection involves us, the value of connecting to correct values just unrolls from the previous causes, so we do not add any value. We do not originate any value.

A daub of paint or a brushstroke in a painting may increase its aesthetic value, but that paint does not bring new value into the

* Compare the sociobiological view that we are simply the arena, the vehicle that genes use to reproduce others like themselves (see R. Dawkins, *The Selfish Gene*, Oxford University Press, 1976), or for that matter Samuel Butler's remark that a chicken is an egg's way of making another egg.

world; the value added in the placement of paint is added by the painter. Yet if he has no more autonomy than a paint blob, then although he is the vehicle through which value is added, and though his act of painting may be a component that is valuable, he does not originate any new value. The process of adding value occurs through his act, but that act does not change or alter the probability (or certainty) that there will be that value in the world. That probability is fixed beforehand and independently of the act, then is played out through the act. The probability of a strike at bowling is not altered by anything the ball does; that probability is fixed by other events: the precise release, condition of the lane, adhesions on the ball and so on. (Nor is the probability altered by the bowler's body movements after the follow-through, when the ball is more than halfway down the lane; follow-throughs in sports are designed to affect, by being headed for, some earlier motions.) The strike occurs through the action of the ball, but the ball's action does not add any new value that (to mix metaphors) was not in the cards already.

We can demarcate four sorts of value something such as an action, thing, event, or state of affairs might have. First, its intrinsic value, the value it has in itself apart from or independently of whatever it leads to or its further consequences. Second, its instrumental value, which is a function and measure of the intrinsic value it will lead to: either the sum of the intrinsic values of the different things it actually will lead to (if these intrinsic values are independent) or some measure of the different intrinsic values it might lead to as weighted by the probabilities that it will, such as the expected intrinsic value. (Or, if there are not even fixed probabilities, a measure appropriate for the uncertainty situation.) Something of instrumental value need not be valuable in itself, it needn't have its own intrinsic value; its instrumental value is measured by the value to which it leads. (We can let considerations of overall theoretical simplicity decide whether something of intrinsic value also has at least its own weight in instrumental value since the probability that it occurs, given that it occurs, is one.) These first two types of value are familiar from the literature.

Third, there is something's originaive value which is a function of the value it newly introduces into the world, the new instrumental or intrinsic value it introduces that was not presaged by or already fully counted in previous instrumental value. An intermediate stage of a

causal process has instrumental value in that it leads to some later intrinsic value, but it lacks origivative value; the probability distribution over future intrinsic value was exactly the same beforehand and it is unchanged by the occurrence of this causally intermediate event. Anything may have all three kinds of values combined, intrinsic, instrumental, and origivative. Note that a randomly occurring event of instrumental value, leading to further intrinsic value, can have origivative value, as this notion was just explained. Perhaps we want to formulate a still tighter notion; but causal determinism denies us even this origivative value. Puppets and marionettes lack origivative value (except in fairy stories), and the way we resemble them, if causal determinism is true, is that we lack origivative value too.²⁹

A being with origivative value, one whose acts have origivative value, can make a difference. Due to his actions, different value consequences occur in the world than otherwise would; these were not in the cards already (with the person's action being one of the cards). It is clear, I think, that a proponent of free will means the choice that is up to us is among actions differing significantly in value. She does not think we can do simply anything at all; our (current) range of choice is limited, for example, to actions that don't involve travel at faster than the speed of light. But neither would she be happy to be told that someone did have a nonunit range of choice but that none of the actions in that range differed significantly. If a murderer was free to choose between stabbing with a knife in one hand and stabbing with it in the other, or to choose between stabbing and shooting, but all the actions he could choose among involved murder of some sort or another, then this would not be a freedom of decision worth having. What is wanted is that we be free to choose among actions that differ significantly in value, or at least in value profile—in the kinds of values they realize, if not in total value score. We want our decisions to make a value difference. Actions of instrumental nonorigivative value do make some difference too when, if they weren't to occur, the later intrinsic values wouldn't materialize. But although such actions make a difference in that they participate in one, they are the vehicles of one, they do not start any difference off, they do not originate one. A paintbrush or a palate knife used by a great painter contributes to some further effects, it is the instrument

whereby they are effected but, unlike the painter, it does not originate any. The palate knife may have its own peculiarities, though, so that the same effects would not occur if another were used. The probability distribution over future effects then must take account of features of the palate knife. (Or need it consider only the factors that give rise to those features?)

Another notion of what difference something makes focuses upon what would have been the case if it had not existed or occurred, but other factors were reorganized to minimize its absence. Here, its value contribution is subjunctively defined by the difference between the actual situation of its presence and the reorganized one in its absence, not by what it leads to causally as a vehicle. Something can be necessary for the production of a certain value, even if it is not an origivative factor. One thing we want is that our actions make a value difference, not merely that value be produced via them but that they be necessary for that production: the slack of their absence would not be taken up by other factors so that the same value got produced anyway. Let us term this value difference something makes—the value (amount) that wouldn't be there if it weren't—its contributory value. We certainly want our actions to have contributory value.

Causal determinism sometimes is misread as fatalism, as saying that our actions have no contributory value because the same (future) thing will happen anyway no matter how we act.³⁰ However, causal determination does allow contributory value; without our actions the future would be different. What causal determinism does not allow is origivative value. The worry about determinism, I think, is that it leaves us no origivative value. The "soft determinist" assures us that contributory value is enough. Yet a puppet can have contributory value also if in no other way could the children be brought to laugh so; although that is value indeed, it is not value enough, of the right sort. We want it to be true that in that very same situation we could have done (significantly) otherwise, so that our actions will have origivative value. The philosophical task is to explain how this is possible.

In explaining the kinds of value, we explained intrinsic value first, then instrumental, origivative, and contributory by the different ways these are related to the intrinsic value that follows them. In the

next chapter, we will investigate the special worth of originary value, which does not lie merely in its contribution to other intrinsic value. For now, let us note that something can be instrumental to later originary value, or originate another thing of originary value, as when parents decide to have children. (When you create certain decision opportunities for others who can freely choose whether to utilize them, how is the originary value of your act to be calculated?) One standard position on the problem of evil is that God created beings (people or angels) with free will, and that this had value even though it opened doors to the production of evil; on this view (in our terminology) God created beings of originary value and this made the originary value of his act of creation higher than it otherwise would have been.

The free decision, bestowing weights on reasons of previously indeterminate weight, has originary value; and as we shall see in the next chapter, it has intrinsic value as well. Presumably, what we would most want are decisions of originary value that track bestness, ones connected to bestness (though not causally) as belief is to truth when it constitutes knowledge. If there was a conflict between the originary and contributory values of a decision, which would we favor, what tradeoff would we make? How much originary value would we sacrifice for (how much) caused increase in the amount bestness is tracked—either in the amount which is tracked or in the fidelity with which it is tracked? And would we want this choice about the tradeoffs we make to be an originary choice or simply one caused to be aligned with value, a self-subsumptively originary or a caused tracking choice?

Let us close this section by considering how we naturally tend to express our worry about determinism: if all of our actions are causally determined, eventually going back to causes occurring before we were born, then. . . . Why is that addendum made about causes before we were born, why is it so natural—rather, what function does it serve, what other possibility is it introduced to block or to cut off? Well, it is clear, isn't it, that if the causes go back to a time before we were born, then we don't control them and so, since they control our decision, we don't control our action. But is the notion of control transitive? Even if so, it does not follow that only the first event, and nothing intermediate in the chain, controls the last event. Even if I build, install, and set a thermostat, controlling it and controlling its

controlling the temperature in my house, still, it does control the temperature in the house. (No one has ever announced that, because determinism is true thermostats do not control temperature.) My decision can control an action decided upon, as an instrumental control and as a contributory control—to adapt the terminology of the earlier value distinctions. The phrase “due eventually to causes existing before I was born” is meant to exclude my being in originary control, initiating a new causal chain that was not already in progress. If, on the other hand, the causes of my current action went back only to last week or last year, then I might control those causes; I might have originated them by some earlier act or decision of mine. To ensure that the causal chain leading to my action isn't itself under my control, the point is added about going back to causes before I was born. For surely I do not control those. No decision or action of mine could originate things occurring before I was born.

Thus, mention of causes going back to before I was born is supposed to make plain that my decision (and I too) can have no originary value. On the view we have presented, however, some actions are not causally determined, though they are, we correctly say in retrospect, caused by the reasons upon which the greater weight was bestowed in that very decision; so the causes may go back a long time, even to before birth, yet nonetheless the person still can originate actions. For which action she does, A or B, is under her originary control, and though the occurrence of the reasons for each, R_A and R_B , are not under her control, the fact that one of them causes (though doesn't causally determine) her act—which one does so—is under her control. She can choose A or B; if she chooses A she makes it true that R_A caused A while if she chooses B she makes it true that R_B caused B. The existence of the cause is not under her control and doesn't originate with her, but the fact that *it* causes her act is and does.

This probing of why we so quickly slide to speaking of causes before we were born fits with the view that it is originary value that is crucial to the problem of free will.* It is difficult to feel any confi-

* Yet, I have the feeling there is more to be discovered here, that that natural move to talking of causes before one was born covers and masks a deeper fact (perhaps a nonclosure to match that exhibited by tracking)—one whose delineation has eluded me, thus far.

dence, though, that this Part I has succeeded in delineating a coherent concept of free action, undetermined yet nonrandom, ascribing originary value to us. (This apart from the question of whether that conception, if coherent, actually is realized.) If indeterminism seems like intellectual quicksand, can we do any better on the more solid but frozen ground of determinism?

II. DETERMINISM AND ALIGNING WITH VALUE

Tracking Bestness

There are two reasons to investigate the notion of an action's tracking bestness. First, that is what we want the upshot of the indeterminist free decision process to be; we want the person who makes free decisions to end up tracking bestness. Second, if determinism is true, no person or action will have originary value. What remains as the most valuable possible mode of action, in that case, is (to be determined) to do the action that is the best of those available to us; moreover, that this doing what is best or most valuable not be an accident, like true belief, but rather that the action tracks value or bestness. Thereby, we and our actions would be connected to value by the subjunctive tracking relation, and to that extent exhibit a valuable mode of action; we would be valuable although not originary of value.

Yet determinism threatens even this, apart from any doubts (which we shall investigate in the next chapter) about the objective status of value. Even if there is value, can our actions track it; will causal determination not attune actions to other (nonevaluative) characteristics? Is causal determination compatible with saying, for example, that if the value of the act were different the person would act differently? Our task, therefore, is not only to delineate what an action's tracking value would be like, but also to see whether and how tracking value or bestness is possible given causal determination of actions.

We investigated knowledge in Chapter 3, delineating how belief is therein connected to truth. We found that a person's belief that p is knowledge when it is subjunctively connected, in a specified way, to the fact that p : he knows that p when his belief that p tracks the fact that p . (Leave aside, for the moment, the additional complications of the reference to methods.) A belief's being causally determined does not undercut the desirability of the tracking connection, and might well underlie that connection. Mightn't a mode of action parallel be-