IV*—FREE WILL: FROM NATURE TO ILLUSION

by Saul Smilansky

ABSTRACT   Sir Peter Strawson's 'Freedom and Resentment' was a landmark in the philosophical understanding of the free will problem. Building upon it, I attempt to defend a novel position, which purports to provide, in outline, the next step forward. The position presented is based on the descriptively central and normatively crucial role of illusion in the issue of free will. Illusion, I claim, is the vital but neglected *key* to the free will problem. The proposed position, which may be called 'Illusionism', is shown to follow both from the strengths and from the weaknesses of Strawson's position.

> We have to believe in free will to get along.
> C.P. Snow

Sir Peter Strawson's 'Freedom and Resentment' (1982, first published in 1962) was a landmark in the philosophical understanding of the free will problem. It has been widely influential and subjected to penetrating criticism (e.g. Galen Strawson 1986 Ch.5, Watson 1987, Klein 1990 Ch.6, Russell 1995 Ch.5). Most commentators have seen it as a large step forward over previous positions, but as ultimately unsuccessful. This is where the discussion within this philosophical direction has apparently stopped, which is obviously unsatisfactory. I shall attempt to defend a novel position, which purports to provide, in outline, the next step forward. The position presented is based on the descriptively central and normatively crucial role of illusion in the issue of free will. Illusion, I claim, is the vital but neglected *key* to the free will problem. It is not claimed that we need to induce illusory beliefs concerning free will, or can live with beliefs we fully realise are illusory—both of these positions would be highly implausible. Rather, my claim is that illusory beliefs are in place, and that the role they play is largely positive. The proposed position, which may be called 'Illusionism', can be defended independently from its derivation from Strawson's 'reactive naturalism', but it is helpful to present the progression

in this way. Since the role of illusion emerges only at a late stage of the train of arguments pertaining to free will, we will get to the final destination by 'free-riding' most of the way on Strawson's train, and then continue a bit further by ourselves, into the uncharted and dangerous Land of Illusion.

   This paper consists of six parts. Part 1 sets out reactive naturalism. Part 2 explains why it has been thought to improve on previous positions on free will. Part 3 shows why reactive naturalism is inadequate. Part 4 elaborates on the problems whose solution requires illusion. Part 5 presents Illusionism and motivates it philosophically. Part 6 reviews the road from reactive naturalism to Illusionism.

I

Strawson's 'Humean Naturalism'. Naturalism, in the sense I am concerned with here, is a sort of 'Humean' response to scepticism about our common freewill-related practices and reactions. It will be called here 'reactive-naturalism' and 'naturalism' interchangeably. This sort of naturalism considers scepticism as idle in view of the natural inclinations of humanity, given which there is no need for countering the sceptic or, indeed, for offering any justification at all of our basic beliefs and attitudes (Strawson 1987, 38–41). Like the compatibilist, the naturalist claims that morality and human life are not dramatically affected by the absence of libertarian free will. However, he rests his case not on an analysis of the philosophical implications of this absence but on its insignificance in 'real life'.

   Considering the predominance of human reactive attitudes and their centrality in human life, indeed, in being human, any intellectual considerations, such as the truth of determinism, cannot seriously be posed as a threat. And, even if we can imagine having a choice whether to engage in inter-personal relations, founded as they are on reactive attitudes, rational choice would be based on the expected gains and losses to human life, and the outcome would be clear (Strawson 1982, p.70). Strawson claims that the nature of morality is largely analogical, in that our demands for other persons' good will towards third parties (or demands from ourselves) resemble those we make for good will towards ourselves. This analogy is sustained by the various

excusing conditions, in the particular irrelevance of consider-
ations of the truth of determinism to them, and in the ridiculous
nature of suggestions that the truth of determinism might abolish
this part of life, the moral sphere being intimately connected with
the sphere of inter-personal relations (p.74).

II

The Attractions of Reactive-Naturalism.
II.1. *Preliminaries.* Reactive naturalism is a strange position to
hold on the free will problem, since it is not really about whether
we do or do not *have* free will or moral responsibility in an inde-
pendent sense, in the way that traditional positions (libertarian-
ism, compatibilism and hard determinism) clearly are. Rather,
reactive naturalism focuses on our emotional lives and asks
whether, in the light of our emotional make-up, common views
are liable to be affected. That such a position can be a contender
at all must be due then to a serious state of affairs; this is indeed
the case. This state is, broadly:

   (A) There is no libertarian free will.
   (B) Compatibilism is insufficient as a basis for moral
        responsibility and related matters.
   (C) We need to maintain common freewill-related attitudes
        and practices, so that compatibilist distinctions in terms
        of control and its absence should largely continue to be
        followed.
   (D) Other alternatives, such as utilitarian, are inadequate.

   In considering the free will problem, the first[1] question is
whether libertarian free will really exists, i.e. the libertarian
Coherence/Existence Question. The second question is
whether—if it does not—we are in trouble. It can be called the
Compatibility Question, namely, are moral responsibility and
related notions compatible with determinism (or with the absence

1. The free will problem can be structured in various ways, but the way presented
here seems most useful: if there were libertarian free will, much of the point in asking
the Compatibility Question would disappear, and so on. Nevertheless, one may begin
by asking if compatibilism is insufficient and then move on to see whether libertarian
free will could help. In any case, reactive naturalism emerges at the end of such
traditional explorations and, I claim, Illusionism follows from the weaknesses of reac-
tive naturalism.

of libertarian free will irrespective of determinism)? Compatibil-
ism and hard determinism are the opponents on the Compati-
bility Question. Reactive naturalism, like the Illusionism I am
offering, is best understood as an answer to the third-level ques-
tion of the consequences of pessimistic answers to the first two
questions—namely, that there is no libertarian free will, and that
compatibilism is insufficient and hence we are in trouble. It can
be called the Consequences Question.

It is the despair from the possibility of grounding our ethical
and personal free will-related beliefs, practices and reactions on
libertarianism, compatibilism or any other traditional alterna-
tive, which brought naturalism to the scene.

II.2. *Why Not Libertarian Free Will?* First a few words on (A),
i.e. the claim that there is no libertarian free will. I shall assume,
with Strawson, that libertarian free will is incoherent (Strawson
1980, 265). In a nutshell, the conditions required by an ethically
satisfying sense of libertarian free will, which would give us any-
thing beyond sophisticated formulations of compatibilism, are
self-contradictory, and hence cannot be met. This is so irrespec-
tive of determinism or causality. Attributing moral worth to a
person for her decision or action requires that it follow from
what she is, morally. The decision or action cannot be produced
by a random occurrence and count morally. We might think that
two different decisions or actions can follow from a person, but
which one does, say, a decision to steal or not to steal, again
cannot be random but needs to follow from what she is, morally.
But what a person is, morally, cannot ultimately be under her
control. We might think that such control is possible if she
creates herself, but then it is the early self that creates a later
self, leading to vicious infinite regress. The libertarian project was
worthwhile attempting: it was supposed to allow a deep moral
connection between a given act and the person, and yet not fall
into being merely an unfolding of the arbitrarily given, whether
determined or random. But it is not possible to find any way in
which this can be done.

Libertarians may well not be satisfied with my cursory treat-
ment, but this should be accepted for the sake of the current
discussion, for we need to journey far. We may then say that my
argument, like that of Strawson, is primarily addressed to those

who are not assured of the belief in the existence of libertarian free will.

II.3. *Why Not Compatibilism?* I will now say something on (B), i.e. on why I think that compatibilism, its partial validity notwithstanding, is grimly insufficient. Not only is compatibilism a widely prevalent view in philosophy, and hence I need to combat the complacency it encourages if I am to motivate the need for illusion, but this need will emerge from the situation that makes compatibilism inadequate. (The case made here is my own, and it is not claimed that Strawson would view matters in exactly the same way.)

We can make sense of the notion of autonomy or self-determination on the compatibilist level but, if there is no libertarian free will, no one can be ultimately in control, ultimately responsible, for this self and its determinations. *Everything* that takes place on the compatibilist level, irrespective of the local distinctions in respect of control, becomes on the ultimate hard determinist level 'what was merely *there*', ultimately deriving from causes beyond the control of the participants. If people lack libertarian free will, their identity and actions flow from circumstances beyond their control. To a certain extent, people can change their character, but that which changes or does not change remains itself a result of something, and there is always a situation in which the self-creating person could not have created herself, but was just what she was, as it were, 'given'. Being the sort of person one is, and having the desires and beliefs one has, are ultimately something which one cannot control, which cannot be one's fault, it is one's luck. And one's life, and everything one does, is an unfolding of this. Let us call this the 'ultimate perspective' and contrast it with the 'compatibilist perspective', which takes the person as a 'given' and enquires about her various desires, choices and actions.

Consider the following quotation from a compatibilist:

> The incoherence of the libertarian conception of moral responsibility arises from the fact that it requires not only authorship of the action, but also, in a sense, authorship of one's self, or of one's character. As was shown, this requirement is unintelligible because it leads to an infinite regress. The way out of this regress is *simply to drop* the second-order authorship requirement, which is what has been done here (Vuoso 1987, 1681; my emphasis).

The difficulty is that there is an *ethical basis* for the libertarian
requirement and, even if it cannot be fulfilled, the idea of 'simply
dropping it' masks how *problematic* the result may be in terms
of fairness and justice. The fact remains that if there is no liber-
tarian free will a person being punished *may suffer justly* in com-
patibilist terms for what is ultimately her luck, for what follows
from being what she is—ultimately beyond her control, a state
which she had no real opportunity to alter, hence neither her
responsibility nor her fault. With all the importance of compati-
bilist distinctions, a morally serious compatibilist cannot escape
the conclusion that if this person suffers—however justly in com-
patibilist terms—she is, from an important perspective, a victim.
For it was *given* that being who she was she would (compatibilist-
ically freely) choose as she did, and suffer the consequences.

A similar criticism applies to other moral and non-moral ways
of perceiving and treating people. The compatibilist cannot
maintain the libertarian-based view of moral worth or of the
grounds for respect, and what she has to offer is a much shal-
lower sort of meaning and justification. Desert, be it of praise or
punishment, can make sense only on a shallow compatibilist
level, where the underlying causes of the good or bad motives
are not queried. Ultimately people are not deserving, they are
simply the way they have been made, and hence equal in value,
i.e. equally lacking in desert-based value. Compatibilism, in sum,
is morally, even humanly, shallow, for it depends on our remain-
ing on the level of people as more or less 'givens', i.e. on blindness
as to what we learn when we push our inquiries further, into the
causes of this 'given', beyond the limited internal compatibilist
perspective. The picture of moral reality and of personal aspects
of worth that we can aspire to as compatibilists is often tragic
and inherently shallow. It is those two charges, of a shallowness,
and of a complacent compliance with the injustice of not
acknowledging lack of fairness and desert, and in particular ulti-
mate-level victimisation, which form the backbone of my case
against compatibilism.

II.4. *Why Not Hard Determinism?* If there is no libertarian free
will and if compatibilism is insufficient, should we not then opt
for hard determinism, which denies the reality of free will and
moral responsibility in any sense? I will now briefly defend (C)

above, namely the need to retain some of the 'form of life' based on the value put on distinctions made in terms of compatibilist free will. I share with most free will philosophers the belief in the at least partial validity of compatibilism. In broad outline, the basis for this position combines the reality of distinctions in terms of local free choice even in a world without libertarian free will, such as a deterministic world, and the possibility to motivate ethically the making use of these distinctions. (Again, Strawson would not share this view in its particulars, but since this is not a work of historical interpretation this need not concern us. He *would* certainly share the thought that hard determinism should not guide our practice.)

The kleptomaniac and the alcoholic differ from the common thief and common drinker in the deficiency of their capacity for local reflective control over their actions (see e.g. Glover 1970, 136; cf. Fischer 1994 for a recent sophisticated compatibilist formulation). Here everyone should agree. But the point is that such differences are *morally significant*: in some ways the compatibilistically free may also be victims, as viewing things from the ultimate perspective has helped us to see, and yet the importance of the commonplace compatibilist level distinctions will often be great. Consider, for instance, the notion of a valid Will and Testament made by a person wishing to arrange the distribution of her property after her death. The idea of a valid will requires that it be made of one's 'free will'. There are likely to be borderline difficulties but, in general, we are able to identify what it *is about* the agent and the situation when signing a document which makes the signatory's action free in a sense we care about (even without libertarian free will), and what limitations of free will (such as coercion and insanity) invalidate the will. And it is fairly obvious *why* we want to make use of these factors in our ethical judgements, reactions and social practices. We want our last wishes to be respected, as well as defence if our will is tampered with, and an ethically decent social order will follow the compatibilist distinctions.

More generally, we want to be members of a Community of Responsibility where our choices will determine the moral attitude we receive, with the accompanying possibility of being morally excused when our actions are not within our reflective control, e.g. when they result from a brain tumour. The exceptions and excuses commonly presented by compatibilism should,

in general, continue to carry weight. For if people are to be respected, their nature as purposive agents capable and desirous of choice needs to be catered for. We have to enable people to live as responsible beings in the Community of Responsibility, with their lives based largely on their choices, to note and give them *credit* for their good actions, and to take account of situations in which they *lacked* the abilities, capacities and opportunities to choose freely, and are therefore not responsible in the compatibilist sense. Even without libertarian free will, it is reasonable to desire that compatibilist distinctions concerning control affect the way one is treated, and to see this as a condition for civilised existence.

Such a community is possible on the basis of compatibilist-level distinctions. Except in extreme situations, we have no reason to accept at face value the words and deeds of a woman who admits that she continuously takes advantage of people and treats them shabbily but claims, as an excuse, that 'this is in her nature'. Admittedly, it may be more difficult for her to control herself or to change than it is for others but this can be only a mitigating element, and would not lead us to accept her presentation of things as simply an acceptable excuse. We shall see her self-justification as, at best, bad faith but more probably as an attempt at self-serving manipulation. One is not normally in a *passive* relationship with such features of one's behaviour, and is an agent who deliberates, decides and acts out one's decisions, not a spectator of forces carrying one along. This element of 'up-to-usness' is why the compatibilist perspective is available, why we are allowed to hold this woman accountable and why we are permitted to attempt to influence her within a responsibility-based moral structure. Such a Community of Responsibility allows people to live lives of integrity based upon their choices, and is also a basis for a fair division of burdens. As Will Kymlicka points out: 'It is unjust if people are disadvantaged by inequalities of their circumstances, but it is equally unjust for me to demand that someone else pay for the costs of my choices' (Kymlicka, quoted in Cohen 1989, 933). Hence, with all the moral importance of the absence of libertarian free will, we need not and must not escape from living according to the basic ethical paradigm of control and responsibility.[2]

2. We need to combine the insights of compatibilism and hard determinism into a joint position, for neither on its own is adequate; see Smilansky (1993) and Smilansky (2000, Part I).

We see, then, that we are in serious trouble, for (A) libertarian free will does not exist, (B) compatibilism is greatly insufficient, but (C) the basic free will-related practices and reactions should be maintained.[3] Naturalism attempts to offer, as we have seen, a defence of our common attitudes and practices in this predicament.

III

Why Not Reactive-Naturalism? Reactive-naturalism offers a highly significant contribution to the debate. When examined closely, however, it is much weaker than it seems at first, and if the basic attitude it favours is to be sustained it is necessary to interpret it along the lines of Illusionism.

III.1. *Revisionist Naturalism.* Naturalism can be seen to split into two versions: a tough-minded, revisionist naturalism, and a softer, passive account. Revisionist naturalism seeks to change the perception that there is a theoretical need to justify common attitudes and practices, holding that there is no need for general grounding and that the reactions themselves provide all the (self-) grounding required. Strawson sometimes expresses views akin to revisionist naturalism (e.g. 1987, 32–3). In greater detail, Jonathan Bennett might be interpreted as working towards such a position in his attempt to apply Strawson's work to the issue of punishment (1980b; cf. Wallace 1994). Bennett suggests that we see the reactive attitudes as the limiting addition (constituting 'justice') to the regular consequentialist considerations regarding punishment, and this explains why we should not 'punish' the innocent, for example—we cannot resent them, and to 'punish' the innocent would be harmful to our reactive lives (pp.48–9).

3. Further views, such as utilitarianism, can be thought to be relevant. I cannot consider the general merits of utilitarianism here, but it seems to me to be fundamentally alien to the deep concerns of the free will issue. It is radically at odds with moral phenomenology: the seriousness of moral appraisal depends on our not viewing judgements merely as manipulative ways of influencing people, which can in principle be applied to the blameless if it is socially useful to do so. People would not be willing to be blamed, would not accept blame as appropriate, were it not assumed that they deserve blame on account of their freely taken actions. Utilitarianism is also opposed to our deepest moral intuitions (the inherent concern with control in general and the abhorrence for the 'punishment' of the innocent example in practice). Even if one accepts utilitarianism the role of illusion with respect to free will can be demonstrated, but this paper can be taken to address those whose basic ethical views are not (only) utilitarian. I cannot consider other positions here.

There is likely to be much affinity between the structure of our reactive lives and the basic ethical intuitions requiring the existence of free will. This depends, however, on the reality of distinctions in control, which influence our knowledge when a reaction is *appropriate*. Our reactive attitudes are not independent and self-validating. The concerns of the basic ethical intuition and reactive attitudes may differ and, in any case, the needs of our reactive lives cannot be the main consideration. Justice, for instance, involves matters other than safeguarding the reactive attitudes. If we doubt whether free action exists in a significant sense, this must be crucial to the view we take of the justification of blame and punishment, for example. Even with Strawson's paradigmatic attitude, resentment, a belief *transcending and underlying* the attitude itself seems necessary. As Joel Feinberg says: 'It is clear, I think, that resentment without an ostensible desert basis is not resentment' (Feinberg 1970, 71).

Similarly, the wish to preserve the reactive attitudes could hardly be widely accepted as a basis for morality. Revisionist naturalism can be confronted with the same type of criticism that naturalism itself levelled at consequentialism in the free will context—that, like the consequentialist 'effects of blame', safeguarding the reactive attitudes is just not the *kind* of reason bound up with freewill-related moral life, and will not be recognised by most people as appropriate. Dependency on grounds is inherent in the notions under consideration, such as resentment, blame and punishment. The reactive attitudes follow the existence of such grounds (e.g. gratitude has to be deserved), and cannot in themselves replace it.

III.2. *Non-Revisionist Naturalism.* Perhaps a non-revisionist, passive naturalism is sufficient? Perhaps all the naturalist requires is that *in practice* common attitudes and behaviour remain constant, whatever the theoretical case may be. However, even on its own naturalistic terms, naturalism is inadequate. Even someone such as Paul Russell, who follows Hume and Strawson in discounting the need for general justification of our moral attitudes as such, thinks that Strawson is too optimistic as to the stability of specific attitudes and practices (Russell 1992). But why are reactive attitudes insufficient?

III.3. *Justice.* One problem that has real practical substance concerns justice. It is not required that people think of some alternative moral position, unrelated to the reactive attitudes. It is enough that some segments of the population should become *cynical* or *doubtful*, about the moral difference between the guilty and innocent (in traditional terms) in the light of the causal role of crimogenic environments, in order for any confidence as to the assurance provided by reactive attitudes to be shaken. In such a case we might face a threat to the basis of moral life in the requirement for considering free will as a condition for punishment. Societies with very different conceptions of justice existed, and even if they cannot be reinstated, significant doubts as to the justness of our own institutions, resulting in an uncaring cynicism, cannot be ruled out.

A sense of 'justice' has often been closely connected with feelings of revenge, concern with the existence of free will being non-existent or meagre: it has often been thought that 'just' revenge could be taken on the tribesmen or countrymen of the guilty person, without undue concern for their lack of responsibility. Today such attitudes are still expressed in societies organised largely on the basis of kinship, resulting in blood-feuds. Many terrorist actions, such as blowing up civilian airlines or buses in the name of 'just' revenge, show the same disregard for the value of considering free will. Such beliefs, emotions and practices were long considered natural, and the danger has not been eradicated. Consider also the prevalence for hundreds of years in Christian Europe of the sentiment that Jews, as such, should suffer, because of the alleged role of some Jews in the Roman crucifixion of Jesus. Other religious beliefs, such as the common conception of Original Sin under certain interpretations, might also betray the lack of concern with 'up-to-usness'. Here we are faced with situations where even the *minimal* content of free will—agency— is not considered necessary in order for punishment to be just.[4]

Let us call the ethical demand for considering the existence of free will in justifying e.g. blame and punishment the 'Core

---

4. Natural human inclinations have not always been sufficient to safeguard our core conception. For more historical examples, from the ancient Greeks, medieval societies, and anthropology, see Sayre (1932, 977 and 981); Adkins (1960, 57, 68, 167); Hibbert (1963, 201f.); Von Furer-Haimendorf (1967, 216); Pollock and Maitland (1968 vol.2, 470f.).

Conception' of justice. I believe that current acceptance of the Core Conception, in some societies, has a good claim to represent moral progress, perhaps the best claim there is: think about enlightened attitudes towards the 'punishment' of the innocent or collective 'punishment', for example. This progress is more fragile than we are wont to think. Moreover, such threats need not even arise from within the free will issue. There are various intellectual and social currents which might harm the value put upon free will. The call for 'efficiency' in the fight against crime might suffice here. The point is that even a mild weakening of free will beliefs might reduce their power to make us *resist* such external influences.

III.4. *Respect*. We can highlight many of these points by considering the issue of (self-)respect. Just as with the issue of justice, I think consideration of this topic shows Strawson's position to be too optimistic. This becomes apparent if we substitute 'respect' for 'resentment' and consider the matter of 'Freedom and Respect', respect being an attitude much more dependent on complex cognitive beliefs than resentment. Resentment may linger when we cease to see the issue of (dis)respect as pertaining to a person, in light of the issue of free will. We can hardly continue to respect ourselves in the same way if we really internalise the belief that all action and achievement is ultimately down to luck and not ultimately attributable to us. And there is every reason to believe that many educated people can internalise this thought to a degree that will *suffice* to cause serious harm to their self-respect. Similarly with the appreciation of and respect for others.

Moreover, the non-revisionist naturalist position, to the extent that it is deemed convincing, is itself harmful to our self-respect. If attitudes of respect are thought unjustifiable on the deepest level because of the lack of an ultimate basis for them in free will, then to say that it is 'unavoidable' that we hold such attitudes means that we are caught in a humiliating state. To continue with the same attitudes and practices involving libertarian desert and worth when we know that there is no libertarian free will is hardly conducive to self-respect, even if no real choice is involved.

In sum, the issue of (self-)respect illuminates the danger of the ultimate hard determinist insight, and shows the weakness of a

position such as Strawson's. Firstly, we realise that what we seek is a deep basis for being worthy of respect, and this Strawson cannot give us. The quest for respect is a quest for true appreciation and value, and cannot be satisfied in the way that the need for some of the more 'emotional' reactive attitudes perhaps can. Secondly, we see that Strawson relies too heavily on our natural proneness to the reactive attitudes as a means of upholding these values, since many people may come to doubt the basis for their self-respect. Finally, even where they are effective, Strawson's pragmatic assurances are only comforting at the price of a further reduction in our self-respect, due to their very nature, and because of our very need for what might be taken to be mere palliatives.[5]

III.5. *Why Not Reactive-Naturalism? Conclusion.* Reactive-naturalism originally appeared as a strong position for, unlike other compatibilist stances, it in a way encompassed the common libertarian assumptions, while neutralising them, so that they seemed to result, in practice, in compatibilist conclusions. But as we saw, this crucially depends on assuming that the reactive attitudes guarantee the status quo, with the cognitive status of the assumptions and resulting actions following the reactive attitudes. And this is unconvincing.

Free will and moral responsibility are the stuff of our beliefs and convictions and not mere secretions of our natural reactions. Instinctive nature is not in complete mastery in matters of free will, and our personal and ethical convictions can be led up grim paths if the absence of libertarian free will is internalised, as it may be, in part. Our ethical common-sense is not 'built-in', and even those reactive attitudes that are more or less unavoidable cannot guarantee it. Reactive naturalism is a useful antidote to extreme cognitivism concerning free will, but in it the pendulum has swung too far the other way. Despite the role of the reactive attitudes, the free will problem can be important in practice, perhaps mainly in less subjective and moral areas, because we can have limited threats to the central values involved, threats that are perhaps largely unavoidable in the modern world, and should

5. I consider the issue of free will and self-respect in greater detail in Smilansky (1997) and Smilansky (2000, section 6.4).

not be avoided by abandoning the Core Conception values, even if they could be.

<div align="center">IV</div>

Elaborating '*The Problem*'. In order to see how illusion is crucial, we must deepen our understanding of the difficulties which (would) prevail without it. There are fatal weaknesses in naturalism as a solution; but why is there an urgent problem in the first place? We have already seen some difficulties. In what follows I will give a number of further illustrations.

IV.1. *The Question of Innocence*. The danger concerning respect for moral innocence was mentioned above. Even in a world without libertarian free will, the idea that only those who deserve to be punished in light of their free actions may be punished is a condition for any civilised moral order (cf. Hart 1970). 'Punishment' of those who did not perform the act for which they are 'punished', or did so act but lacked control over their action in any sense, is the paradigm of injustice. Yet while the justification for these values does not require libertarian free will, in practice they might be at risk were the lack of libertarian free will internalised. Consider Anscombe's passionate remark that 'If someone really thinks, *in advance*, that it is open to question whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration—I do not want to argue with him; he shows a corrupt mind' (Anscombe 1981, 40). Surely, if a moral system that seeks to preserve and guard vigilantly the common conception of innocence is to function well, such a sentiment should be prevalent, almost instinctive. But if this is to be so, the worst thing one could do would be to point out that, ultimately, none of this makes sense— because the 'guilty' are, ultimately, no more guilty than others. In a world imbued with a deterministic outlook the ethical-emotional weight of the Dreyfus affair, for example, is scarcely comprehensible.

IV.2. *The Ultimate Conclusion as a Practical Threat to the Taking of Responsibility*. We cannot tell people that they must behave in a certain way, that it is morally crucial that they do so, but then, if they do not, turn and say that this is (in every case) excusable,

given whatever hereditary and environmental influences have operated in their formation. Psychologically, the attribution of responsibility to people so that they may be said to justly deserve gain or loss for their actions requires (even *after* the act) the absence of the notion that the act is an unavoidable outcome of the way things were, is ultimately beyond anyone's control. Morality has a crucial interest in confronting what can be called the *Present Danger of the Future Retrospective Excuse*, and in restricting the influence of the ultimate hard determinist level. To put it bluntly: people as a rule ought not to be fully aware of the ultimate inevitability of what they have done, for this will affect the way in which they hold themselves responsible. The knowledge that such an escape from responsibility, based on retrospective ultimate judgement, will be available in the *future* is likely to affect the *present* view, and hence cannot be fully admitted even in its *retrospective* form. We often want a person to blame himself, feel guilty, and even see that he deserves to be punished. Such a person is not likely to do all this if he internalises the ultimate perspective, according to which in the actual world nothing else could in fact have occurred, he could not strictly have done anything else except what he did do.

IV.3. *Failure*. It might also be interesting to reflect upon *failure*. The threat of failure is central to the widespread motivation to study, work, and in general make an effort, i.e. in motivating achievement. The sense of achievement and the self-respect it generates are in everybody's interest; unfortunately these ideas make no sense without the notion of failure. Hence we need the idea of failure in order to be given the opportunity to succeed. By now, however, it will be obvious that the ultimate perspective poses a great threat here. If the boy at fifteen is to make something of himself, it cannot be the case that, were he to fail, at sixty he would have an easy way of dismissing his plight as all along beyond his control, for hard determinist reasons. Moreover, such an easy erasure of failure cannot but affect the fate of the sense of achievement: it cannot be that failure is thought not to be in the end up to one, while attainment miraculously remains so. A cultural climate of guaranteed excuse is not conducive to effort and for encouraging success, nor is it a firm foundation for (self-)respect.

IV.4. *A Sense of Value*. From the ultimate hard determinist perspective, all people—whatever their efforts and sacrifices—are morally equal: i.e. there cannot be any means of generating 'real' moral value. As we have seen, there is a sense in which our notion of moral self-respect, which is intimately connected with our view of our choices, actions and achievements, withers when we accept the ultimate perspective. From the latter any sense of moral achievement disappears, as even the actions of the 'moral hero' are simply an unfolding of what he happens to be *no matter how devoted he has been, how much effort he has put in, how many tears he has shed, how many sacrifices he has willingly suffered*. True *appreciation*, deeply *attributing* matters to someone in a sense that will make him worthy, is impossible if we regard him and his efforts as merely determined products. All that the compatibilist can offer us in terms of value, although important in itself, is meagre protection from the cold wind that attacks us when we come close to reaching the luck-imbued ultimate level. There is an obvious practical danger here to our moral motivation, which can be named the *Danger of Worthlessness*. But the concern is not only to get people to function adequately as moral agents, but with the very meaning we can find in our lives.

IV.5. *Remorse and Integrity*. If a person takes the ultimate hard determinist perspective, it is not only others who seem to disappear as moral agents—but in some way the person herself is reduced. In retrospect her life, her decisions, that which is most truly her own, appear to be accidental phenomena of which she is the mere *vehicle*, and to feel moral remorse for any of it, by way of truly *owning up* to it, seems in some deep sense to be misguided. Feelings of remorse are inherently tied to the person's self-perception as a morally responsible agent (see Taylor 1985, 107).

It sharpens our focus not to dwell upon those happy to escape accountability, but rather upon those who have good will. Here we confront what can be termed the *Danger of Retrospective Dissociation*, the difficulty of feeling truly responsible after action. One can surrender the right to make use of the 'ultimate level excuse' for normative reasons, and yet perhaps not be able to hold oneself truly responsible (e.g. to engage in remorse), if one has no grain of belief in something like libertarian free will. One

can, after all, accept responsibility for matters that were not up to one in any sense, such as for the actions of others, for normative reasons. But here we are dealing with a different matter: not with the acceptance of responsibility in the sense of 'willingness to pay', but rather with feeling *compunction*. Compunction seems conceptually problematic and psychologically dubious when it concerns actions that, it is understood, ultimately one could not in fact help doing. But such genuine feelings of responsibility (and not mere acceptance of it) are crucial for being responsible selves! We see here the *intimacy* of the connection between moral and personal integrity and beliefs about free will; hence the danger of realising the truth also looms large.

Here the common person's incompatibilist intuitions, for all of their vagueness and crudeness, have captured something that has escaped philosophical compatibilists. Once this larger view of the need to have workable beliefs and sustaining self-images is taken we can no longer contemplate with equanimity the decline of libertarian-based beliefs. When we appreciate that it is not merely 'external' or 'theoretical' conclusions which may emerge, but that internalised beliefs regarding the free will problem could enter into our retrospective beliefs about *ourselves*, we see that the difficulties caused by the absence of ultimate-level grounding are likely to be great, generating acute psychological discomfort for many people and threatening morality—if, that is, we do not have illusion at our disposal.

IV.6. *The 'Problem': Some Concluding Reflections.* The difficulties we have seen can be divided into two types. Firstly, reactions and practices which are at least partially valid (have compatibilist grounding) will not be sufficiently adhered to if the absence of libertarian free will is realised. The compatibilist categories are not erased by the absence of libertarian free will, but *over-reaction* to this absence may in practice occur. Secondly, the absence of libertarian free will is *in itself* grimly significant, hence its realisation is potentially problematic irrespective of the danger to the compatibilistically-valid reactions and practices. Even if people continue to respect the compatibilist categories, they may come to see that the lack of libertarian free will is, say, corrosive of their self-respect. As we shall see shortly, illusion assists us with these two problems.

Belief can be fairly stable concerning libertarian free will but, if this current stability-point is broken, it is not the partially valid compatibilist categories that will be upheld. Rather, the risk is that belief will collapse to its next 'natural' stability-point, to the denial of meaning to free will and moral responsibility: as it were, '*If all is determined, everything is permitted.*' And even when this is not the case, the poverty of the best that the compatibilist has to offer in terms of worth and desert is disheartening, and this grim situation can be realised to some extent.

In theory, alternatives to the concern with free will also present themselves: for example, a purely aesthetic view of life that does not treat achievements as reflecting on a person's value, except for a merely quasi-aesthetic ranking. Such abandonment of value and of self is at best a marginal possibility, at least within the framework of anything resembling Western forms of thought. Note that this extends beyond those with deep moral concerns. A true understanding of what is at stake concerning non-moral self-respect, for example, would lead one to the same conclusion. *There is no real substitute for the framework of achievement, desert and value based on free action.* And within that framework, a deep view not diverted by illusion will find itself face-to-face with darkness.

V

Illusion As A 'Solution'.

V.1. *What Is Illusionism?* Illusionism is the position that illusion often has a *large and positive* role to play in the issue of free will. In arguing for the importance of illusion I claim that we can see why it is useful, that it is a reality, and that by and large it ought to continue. As I noted above, it is not claimed that we need to induce illusory beliefs concerning free will, or can live with beliefs we fully realise are illusory. Rather, my claim is that illusory beliefs are in place, and that the role they play is largely positive. Humanity is fortunately deceived on the free will issue, and this seems to be a condition of civilised morality and personal value.

The importance of illusion flows from the basic structure of the free will problem that we have seen. It flows in two ways: first, indirectly, from the fundamental dualism on the Compatibility Question—the partial and varying validity of *both*

compatibilism and hard determinism. The partial validity of compatibilism does not reduce the need for illusion so much as it complicates it and adds to it, because of the need to guard the compatibilist concerns and distinctions, and the contrast and dissonance with the ultimate hard determinist perspective. Secondly, illusion flows directly and more deeply—from the meaning of the very absence of the sort of grounding that libertarian free will was thought to provide. We cannot live adequately with the dissonance of the two valid sides of the fundamental dualism, nor with a complete awareness of the deep significance of the absence of libertarian free will. We have to face the fact that there are basic beliefs that morally ought not to be abandoned, although they might destroy each other, or are even partly based on incoherent conceptions. At least for most people, these beliefs are potentially in need of motivated mediation and defence by illusion, ranging from wishful thinking to self-deception.

The sense of 'illusion' that I am using combines the falsity of the belief with some motivated role in forming and maintaining that belief, as in standard cases of wishful thinking or self-deception. However, it suffices that the beliefs are false and that this conclusion would be resisted were a challenge to arise; it is not necessary for us to determine the current level of illusion concerning free will.

V.2. *Why Is there a Need For Illusion?* Our previous results supply the resources for an answer. Let us concentrate, for the sake of simplicity, on the concerns of a strictly 'practical' point of view: if the basic ethical concern for free will, the Core Conception, is taken seriously, while the absence of libertarian free will is to some extent realised, and illusion does not prevail, then the ultimate level conclusion might tend to dominate in practice. It might very well pose a danger—especially because of the human tendency to over-simplify—to the 'common form of life' and to the strict observance of the corresponding moral order. Many people would find it hard to think that the partial compatibilist truth *matters*, as in fact it ethically does, if they realised the sense in which both the compatibilistically free and the unfree were merely performing according to their mould. And this might lead them to succumb to 'pragmatic' consequentialist temptations, or an unprincipled nihilism. The ultimate hard determinist perspective does not leave sufficient moral and psychological 'space' for

compatibilistically-defensible reactive attitudes and moral order. The fragile compatibilist-level plants need to be defended from the chill of the ultimate perspective in the hothouse of illusion. *Only if we do not see people from the ultimate perspective can we live in a way which compatibilism affirms*—blaming, selectively excusing, respecting, being grateful, and the like.

Within these parameters, there is a *prima facie* case for a large measure of motivated obscurity regarding the objections to libertarian free will: if libertarian assumptions *carry on their back* the compatibilist distinctions, which would not be adhered to sufficiently without them, an illusion which defends these libertarian assumptions seems to be just what we need.[6] The partial validity of the compatibilist distinctions is unlikely to overcome the practical salience of the ultimate perspective in such a situation, unless illusion intervenes. Determinists are not likely to cherish and maintain adequately the respect due to people in the light of their free actions, nor a free will-based moral order in general. The ethical importance of the paradigm of free will and responsibility as a basis for desert should be taken very seriously, but the ultimate perspective threatens to *present* it as a farce, a mere game without foundation. Likewise with the crucial idea of a personal sense of value and appreciation that can be gained through our free actions: this is unlikely to be adequately maintained by individuals in their self-estimates, nor warmly and consistently projected by society. A broad loss of moral and personal confidence can be expected. The idea of action-based desert, true internal acceptance of responsibility, respect for effort and achievement, deep ethical appreciation, excusing the innocent— all these and more are threatened by the 'levelling' or homogenising view arising from the ultimate perspective. Illusion is crucial in pragmatically safeguarding the compatibilistically-defensible elements of the 'common form of life'. *Illusion is, by and large, a condition for the actual creation and maintenance of adequate moral and personal reality*.

V.3. *How Does Illusion Function?* When illusion plays a role, things can, in practice, work out. Two schematic answers can be

---

6. There are many complex ways in which illusion may be functional concerning free will, which we cannot consider here (see Smilansky 2000, section 8.4). We have focused on the main way, helping maintain false libertarian beliefs.

made. Significant realisation of the absence of libertarian free-will and concern about ultimate level injustice, for example, can remain more or less limited to part of the population, say, those more concerned with policy-making (an 'elitist solution'). This maintains the widespread 'intuition' that, for instance, 'punishing' the innocent is an abomination whereas criminals deserve 'to pay', while permitting the amelioration of treatment, resulting from the recognition, by some, that ultimately things are not morally that simple. Complex patterns of self-and-other deception emerge here. But, in addition to all the general practical and moral difficulties with elitist solutions, which we cannot consider here, elitism can in any case be only a partial solution concerning freewill. For, in the light of the reasons that we have already seen, people without illusions would have great difficulty in functioning.

The major solution will be one where, since two beliefs are vaguely but simultaneously held, yet commonly not set side by side (often, I claim, due to the presence of a motivated element), their contrary nature is not fully noticed. When acting in the light of compatibilist insights we suspend the insights of the ultimate hard determinist perspective (which we in any case are likely to be only dimly aware of). We *keep ourselves* on the level of compatibilist distinctions about local control, and do not ask ourselves about the deeper question of the 'givenness' of our choosing self; resisting threats to our vague, tacit libertarian assumptions. As Bernard Williams put it: 'To the extent that the institution of blame works coherently, it does so because it attempts less than morality would like it to do... [it] takes the agent together with his character, and does not raise questions about his freedom to have chosen some other character' (1985, 194). The result is not philosophically neat, but that, after all, is its merit: the original reality was that we face practical dangers if we try to make our (incoherent or contradictory) conceptions *too clear*, but that we ought not to give any of them up entirely. Illusion, in short, allows us to have 'workable beliefs'.

We can *expect* people to be able to function adequately when they are compatibilistically free. There is ample basis in compatibilist local control for doing so, and enlisting such functioning is a condition of civilisation. When we remain on the compatibilist level, distinctions and excuses emerge which allow for normal

human interaction, for our reactive lives, for the accumulation of moral credit and discredit and for moral discernment. However, awareness of the ultimate inevitability of any level of functioning endangers *good* functioning, and *darkens* our fundamental ways of appreciating ourselves as well as others; hence illusion is required. Illusion not only functions in motivated resistance to threats to our beliefs; but it also offers a positive view *underlying* our attitudes and practices. The affirmation of the responsible self is furthered by the vague tacit belief that one was and is able to do otherwise in the libertarian sense, and can have no general escape from the burden of responsibility. It is not that we find out the truth and then say 'Let's keep quiet about this', but that illusion is intimately *entangled* with our freewill-related beliefs, reactions and practices. However, some awareness of deterministic elements can be useful, mitigating resentment of others or self-recriminations. Illusion allows us the advantages of the libertarian picture together with the mitigating element, without full awareness either of the incoherence of the libertarian picture or of the contrariness of the compatibilist and ultimate perspectives.

The interaction between illusion and reality is subtle: illusion is often the handmaiden of reality and, indeed, its constant support. Matters such as the acceptance of personal responsibility, adherence to the values and practices of a Community of Responsibility, and the sense of pride at having done 'all that one could', are of immense value and find some grounding on the compatibilist level. They can be a non-illusory reality. However, they often depend upon lack of awareness of the ultimate perspective: illusion does not turn everything into falsehood but, on the contrary, is often the condition for the *emergence* of a valid and morally necessary reality.

Moreover, even those elements of our self-understanding that are solely illusory (and not compatibilistically-grounded reality merely assisted by illusion) may nevertheless be very important in themselves. Illusion not only helps to sustain independent reality, but is also *in itself* a sort of 'reality', simply by virtue of its existence. The falseness of beliefs does not negate the fact that they exist for the believer. This is the way in which the libertarian beliefs exist. In addition to supporting the compatibilist non-illusory basis, illusion also *creates* a mental reality, such as a

particular sense of worth, appreciation and moral depth associ-
ated with belief in libertarian free will, which would not exist
without it. The effects of this illusory 'reality' are sometimes posi-
tive. In a number of ways, then, illusion serves a crucial *creative*
function, which is a basis for social morality and personal self-
appreciation, in support of the compatibilist forms and beyond
them.

The idea of illusion as morally necessary is repugnant and
demeaning. As David Wiggins aptly put it:

> If a dilemma exists here it should first be acknowledged and felt
> as such. Only barbarism and reaction can benefit by concealment.
> If the unreformed notion of responsibility, the notion which is our
> notion, is a sort of metaphysical joke must we not at the very least
> create some safe time or place in everyday life to laugh at it? (Wig-
> gins 1973, 55).

Nevertheless, I do not see any resources left to combat the ethical
necessity of illusion in the free will case.

VI

From Naturalism To Illusionism. Revisionist naturalism sought
to neutralise problems such as we saw by saying that we need
not care, that there is no need to justify our attitudes and prac-
tices. Non-revisionist naturalism was more modest, merely
insisting that in practice not much can change, and that for this
reason there is little room for concern. Both stances were found
to be unconvincing. We can understand the 'conservative'
instinct of naturalism but see that illusion is required. The
insights of naturalism can be better defended in combination
with an illusionistic element. We end up with the broad con-
clusion that our priority should be to live with the assumption
of libertarian free will although there is no basis for this other
than our very need to live with this assumption; but as we cannot
accept this way of seeing things, and confront dangers to our
beliefs, illusion must play a central role in our lives.

Reactive naturalism is important for Illusionism. Firstly, the
failure of naturalism's 'don't-worry' attitude leads to Illusionism.
Naturalism has been seen by many as the last hope of compatibil-
ism, and its weaknesses lead to recognition of the *role* and the
need for illusion. Those who came to naturalism out of despair

of previous alternatives now need to take a further step towards Illusionism. It can be said that naturalism and Illusionism are the last competitors. It is not by chance that they are located at the stage of the third question, the Consequences Question, unlike the more traditional positions. The progression to the third question reflects our acknowledgement that the answer to the Compatibility Question (are moral responsibility and the associated notions compatible with the absence of libertarian free will?), for all its importance, is not conclusive, and, moreover, that the deep meaning and practical significance of the free will issue is not fully encapsulated in the absence of libertarian free will and the answer to the Compatibility Question. The move to the third question (which asks about the consequences following from our previous results), and the insufficiency of naturalism on that question, serve to firmly identify illusion as the deep factor in the free will issue.

Finally, even more crucially, naturalism indicates the basis for illusion's practical *actualisation*. Naturalism's partial success, and not only its limitations, are instructive here. We are 'naturally' tacit libertarians, and 'naturally' resist threats to free will-related beliefs, attitudes and practices: even when the defence of not seeing threats to libertarian free will in the first place is breached, the damage can be contained. The naturalistic foundation not only paves the way for illusion, but sets it at the heart of the human condition. Illusion is not some external, pragmatic, temporary way of coping with philosophical conclusions, but the very way humanity lives.[7]

Bibliography

Adkins, Arthur W. H., 1960, *Merit and Responsibility* (Clarendon Press, Oxford).
Anscombe, G. E. M., 1981, 'Modern Moral Philosophy', *Collected Philosophical Papers*, vol. 3 (Blackwells, Oxford).
Bennett, Jonathan, 1980, 'Towards a Theory of Punishment', *Philosophic Exchange*, vol. 3, pp. 43–54.
Feinberg, Joel, 1970, 'Justice and Personal Desert', *Doing and Deserving* (Princeton University Press, Princeton).

Fischer, John Martin, 1994, *The Metaphysics of Free Will* (Blackwells, Oxford).

Von Furer-Haimendorf, Christoph, 1967, *Morals and Merit: A Study of Values and Social Controls in South Asian Societies* (Weidenfeld and Nicolson, London).

Glover, Jonathan, 1970, *Responsibility* (Routledge and Kegan Paul, London).

Hart, H. L. A., 1970, *Punishment and Responsibility* (Clarendon Press, Oxford).

Hibbert, Christopher, 1963, *The Roots of Evil* (Weidenfeld and Nicolson, London).

Klein, Martha, 1990, *Determinism*, *Blameworthiness and Deprivation* (Oxford University Press, Oxford).

Kymlicka, Will, quoted in G.A. Cohen, 1989, 'On the Currency of Egalitarian Justice', *Ethics*, vol. 99, pp. 906–944.

Pollock, Frederick and Frederic William Maitland, 1968, *The History of English Law*, 2nd. ed., vol. 2 (Cambridge University Press, Cambridge).

Russell, Paul, 1992, 'Strawson's Way of Naturalizing Responsibility', *Ethics*, vol. 102, pp. 287–302.

Russell, Paul, 1995, *Freedom and Moral Sentiment* (Oxford University Press, New York).

Sayre, Francis Bowes, 1932, 'Mens Rea', *Harvard Law Review*, vol. 45, pp. 974–1026.

Snow, C. P., 1983, in John Halperin *C.P. Snow: An Oral Biography* (Harvester, Brighton).

Smilansky, Saul, 1993, 'Does the Free Will Debate Rest on a Mistake?', *Philosophical Papers*, vol. 22, pp. 173–188.

Smilansky, Saul, 1997, 'Can a Determinist Respect Herself?', in C. H. Manekin and M. Kellner (eds) *Freedom and Moral Responsibility: General and Jewish Perspectives* (University of Maryland Press, College Park), pp. 85–98.

Smilansky, Saul, 2000, *Free Will and Illusion* (Oxford University Press, Oxford).

Strawson, Galen, 1986, *Freedom and Belief* (Oxford University Press, Oxford).

Strawson, P. F., 1980, 'P. F. Strawson Replies' in Zak Van Straaten (ed.), *Philosophical Subjects* (Clarendon Press, Oxford).

Strawson, P. F., 1982, 'Freedom and Resentment' in Gary Watson (ed.), *Free Will* (Oxford University Press, Oxford).

Strawson, P. F., 1987, *Skepticism and Naturalism* (Methuen, London).

Taylor, Gabriele, 1985, *Pride, Shame, and Guilt* (Clarendon Press, Oxford).

Vuoso, George, 1987, 'Background, Responsibility and Excuse', *Yale Law Journal*, vol. 96, pp. 1661–1686.

Wallace, R. Jay, 1994, *Responsibility and the Moral Sentiments* (Harvard University Press, Cambridge).

Watson, Gary, 1987, 'Responsibility and the Limits of Evil: Variations On a Strawsonian Theme' in Ferdinand Schoeman (ed.), *Responsibility*, *Character and the Emotions* (Cambridge University Press, Cambridge).

Wiggins, David, 1973, 'Towards a Reasonable Libertarianism', in Ted Honderich (ed.), *Essays On Actions and Events* (Routledge and Kegan Paul, London).

Williams, Bernard, 1985, *Ethics and the Limits of Philosophy* (Fontana, London).