

Moral Responsibility and the 'Ignorant Scientist'*

by

**John Forge,
School of Science,
Griffith University**

*I am very grateful to Julian Lamont for this helpful comments on an earlier version.

Suppose someone, call her P, does something which leads via some causal process to an outcome g which P neither intends nor foresees. Could P be *responsible* for g? P is of course *causally* responsible for g, for that is the hypothesis, but could she be responsible in some other sense such that she could be praised or blamed for bringing about g? There are various sorts of responsibility which attach praise or blame, including legal, social and moral responsibility. There is nothing praiseworthy or blameworthy in any of these weighty senses about doing something trivial, so we will assume that g is significant in that it affects people in some non-trivial way. It would, however, appear at first sight that P could not possibly be responsible for something if she neither intends it *nor* foresees it.

Writing in a different context, in connection with science policy and his battle with 'Bernalism', Michael Polanyi claimed that pure research was "doubly unpredictable" with regard to its outside applications. Being unpredictable is evidently not the same as being unforeseen, but it would appear to be a sufficient condition.¹ Polanyi's claim is worth considering if we are interested in assessing the moral (or social) responsibility of the pure researcher for the outside applications of her work, applications which are likely to affect others. I will refer to a scientist who does not foresee such applications as "ignorant" and ask whether ignorance is an 'excuse' in the sense that it removes all responsibility for affects of that work on others. Polanyi's claim seems to imply that pure researchers are always ignorant. I think this is an important issue, given the impact of the results of (pure) science on our lives. Scientists may believe that this is all up to the industrialists, the state, the military and so on, and that they themselves have no responsibility. I want to challenge that belief.

In order to address this issue, it will be necessary to treat certain related matters in a rough and ready way and to neglect others altogether - for instance I shall say nothing about responsibility about omissions as opposed to acts and assume that responsibility only attaches to actions. This is (of course) because the question of responsibility and the scientist is a complex one, involving a number of dimensions. It is, however, necessary to have a framework for discussion in the form of some view of what it is for a scientist (or person in general) to perform an action.² So I begin with some preliminaries, move on to discuss the matter of pure research and foreseeability and then see finally whether ignorance is an excuse. One interesting outcome is that the usual 'symmetry' between praise and blame is broken: it turns out that being ignorant may be grounds for blame, but it cannot be grounds for praise.

PRELIMINARIES

Acting

The class of 'actings' that is relevant herein the first place is the class of *intentional* actings - *actions* in the normal sense - when someone *decides* to bring something about. In this sense, then, an action is a *choice*: to do this rather than that. But to do what? One account goes like this: P chooses to bring about event e by doing f.³ f is some sequence of 'internal' occurrences constituted by physico-chemical processes beginning in the brain, where the decision is made or the intention formed by an act of will, and ending with some 'outward' effect, such as an utterance or bodily movement. Presumably there must be something 'outward' for there to be an action. If I am able to lower my pulse rate by relaxing and if this is all there is to what I do - if I am not, for example, hooked up to a device that signals my pulse to

followers who will do certain things when it reaches 65 - then this is not action. Now e might be the last occurrence in the sequence f and itself be the intended action: I might just want to say "thanks" and nothing more. Or, more interestingly, e might be remote. The idea, then, is that e is the *end in view* and P does what she thinks will bring e about. Let me stress the point again: e is what P *intends* when she does f; in a moment I will distinguish e from g, which also follows from f but which is *unforeseen* (and unintended).

The account just outlined is obviously a sketch that needs to be filled out. For one thing, the causal aspect, the bringing about of e by P, needs elaboration: indeed, one can ask whether the relation between P and e always involves cause. I believe that it does and that the best way to elaborate the account here is in terms of causal processes. Thus we could say that P initiates a causal process which she believes will result in e, and so express the possibility that e be remote in time and place from f but that f is nevertheless connected to e. Another point is that it must be assumed that P is *competent*, that she is able to make choices from a range of alternatives, to realise that there are better and worse means to these ends and to decide between these. These are not trivial matters, but again they cannot be dwelt on here.

So let us say that P can be responsible for e only if she brings about f and that e is connected to f via a causal process. Note that this is weak and hedged about: it is only a necessary condition for a possibility - that P *can* be responsible. The reason for this lack of commitment is that we do want to prejudge the whole question at this stage and certainly we do not want to set intention as a necessary condition - which indeed it is not: think of reckless and negligent actions.⁴ But we can commit, *pace* the qualification about omissions, that P cannot be morally responsible for something if she is not causally responsible for it.

Responsibility

In adding on to one's preferred account of action an account of responsibility, it is necessary to introduce a several more dimensions. For instance, it is presumably necessary to add a qualification about freedom and coercion, and here again the issues are not always clear cut. For instance, it may be thought that P is responsible for e only if she freely does f. While this is normally the case, it need not be. Suppose P took a stupefying drug, which rendered her more or less incompetent, and then did f. P was not in control when she did f, so she could not be said to have done f freely. Nevertheless, she still could be responsible for e. Zimmermann distinguishes between direct and indirect responsibility and grounds this on distinctions between direct and indirect freedom and control (Zimmermann 1997: 414). P is then said to be (indirectly) culpable for e only if she is (directly) culpable for taking the drug, and the latter holds only if she freely took the drug. Now for responsibility for e to be ascribed to P, there must be the right kind of connection between e and taking the drug. For example if P knew she was likely to do f when she took the drug, but not otherwise, then she would be culpable if she did so in the expectation that it would steel her to do f, raise the chances of her doing f, etc. There is much else that a complete account of responsibility would need to include, but we must focus on the issue at hand, which can now be raised in view of these remarks about direct and indirect blame. Let us say that (direct) responsibility attaches when P, by doing f, *knowingly, intentionally and freely does right or wrong*, namely (by bringing about) e by doing f, and thus is praiseworthy or blameworthy. Without further qualification or distinctions, such as those of

Zimmermann, P would not be responsible for g if any of the antecedents were not satisfied. So, if P did not know f would *also* entail g - both e *and* g eventuate from f, and we now focus on g - she could not be held accountable.

If, on the other hand, it could be shown that she was (directly) responsible for being ignorant (of the relation between f and g) - if ignorance were no excuse - then it is possible for her to be (indirectly) responsible for g. And if this were a matter for blame, then the distinction between direct and indirect culpability would, *mutatis mutandis*, apply here.⁵ There is evidently no other alternative: if P is not responsible for being ignorant, then there is no way in which responsibility for g to be attributed to her. This raises some further questions. For instance, does ignorance only give rise to blame, never praise? The answer is yes! More importantly, how can we show that the pure scientist P is responsible for being ignorant?

THE IGNORANT SCIENTIST

The Problem of Ignorance

The term "ignorance" is not meant in any derogatory sense: the scientist is, in the first place, ignorant of the consequences of her work just in case she does not know about them. The issue of real importance is not so much actually not knowing what the consequences of work undertaken now will be in the future, but being able to make some informed judgement about what they might be. I have already referred to the latter as *foreseeability*. The ignorant scientist, for present purposes, is one that has little or no foresight with respect to the consequences of her work. This is vague, but we are dealing here with a general philosophical question and not the practicalities of what it is to foresee an outcome.

Now, it is of course a commonplace - in fact an underlying rationale of Science and Technology Studies - that science has consequences that have shaped the very fabric of society. This has normally taken place via technology based on science, but one should not neglect the impact ideas themselves. There is little doubt that such 'progress' is an admixture of both good and bad elements. I am of the opinion, which I expect is shared by most people, that nerve gases and laser anti-personnel weapons are unqualified evil consequences of science, while therapies for inherited genetic defects like PKU are unqualified goods. Some might disagree: some might hold that weapons of mass destruction are deterrents necessary to keep the peace, while gene therapies amount to tampering with evolution and sustaining unfit individuals that should be allowed to die out. There can be controversy and debate about almost any given case. But the general point remains: science affects us, for good or ill. Given this, there is an obligation to try to promote the good and circumvent the ill. But, again, *can* this be done: indeed, does the 'pure' scientific researcher have any responsibility at all? To answer this question we need some examples.

Pure and Applied Research

Laser weapons, be they anti-personnel or anti-communication, send an energetic and concentrated beam of light over a distance which is significantly longer than that found in laboratory applications. I am not aware of any non-military use of laser light of such magnitude. The weapons were the result of applied research efforts which 'scaled up' laboratory lasers. The point about applied research is that

it is undertaken with a more or less definite end in view. It is research that is *applied* for a certain purpose or *outcome*. Pure research, on the other hand, is said to be research undertaken 'for its own sake', not for any particular outcome (Ronayne 1983, p.35). Thus, the phenomenon of laser light was originally discovered by pure research in 1955; applications quickly followed.

The question of foreseeability and ignorance would seem to be much more easily resolved in the case of applied research. It would appear that a scientist working on an applied research project is in a good position to see the likely outcomes of her work: as we have just observed, to have a given outcome is the very point of applied research. If the sign of the laboratory door says "Good Samaritan Volunteer Gene Therapy for the Poor" or if the security guard at the gate wears a badge with the motto "Beelzebub Industries Weapons Banned by the Geneva Convention Research Facility", then it would be hard to plead ignorance about intended outcomes. Moreover, some scientists choose to go into applied research, especially in the biological and medical fields, because they think that here that they can do good and help others. This is not to say that the outcomes of all applied research is always foreseeable, indeed they may not be, and for several sorts of reasons. But ends are surely more perspicuous than they are for pure research.

It was mentioned above that Michael Polanyi, in "The Republic of Science", claimed that pure research is 'doubly unpredictable' (p.59). What he meant was that the applications of pure research were hard (impossible?) to foresee for two reasons: first of all the purely scientific outcome is not predictable at the beginning of a research project, and secondly, given this outcome, its applications are not predictable. To illustrate the second sort of unpredictability, he told a story of himself and Bertrand Russell being unable to think of any application of the Einstein formula $e = mc^2$ when asked about it in a radio programme in April 1945 (pp. 58-9). This was just three months before the Trinity test of the first atomic bomb, a device designed to convert mass into energy. As to the first sort, presumably Einstein himself did not foresee that this formula would be a consequence of his attempt to find transformations that would restore the symmetry between moving charge and moving conductor. From Einstein's perspective, the bomb was therefore 'doubly unpredictable'.

We should not, however, make the mistake of thinking that the immediate products, the scientific knowledge, of pure and applied research are intrinsically different kinds of things. It is quite possible that the very same piece of work could be classified as either pure or applied, depending on the context. For example, the fission cross-sections of uranium and plutonium isotopes are of interest in and of themselves, pieces of the puzzle that is nuclear physics. But these values are also prerequisites for calculating the geometry of critical assemblies that constitute nuclear weapons. Thus, the Los Alamos scientists working on these problems were doing applied research, but that classification was a function of the context in which they worked.

We can conclude from this that if pure and applied research are distinguished in the usual way, in terms of intentions, then this does not serve to pick out two different kinds of knowledge. If it did, if 'pure scientific knowledge' (to coin a horrid neologism) were in some way inapplicable in principle or had some intrinsically unforeseeable character as regards its outside applications, then pure researchers could not have any responsibility as regards these applications. They would be unforeseeable and ignorance would be an excuse. Since this *not* the case, two

questions arise. There is first what we might call the philosophical question, namely whether the *intention* only to discover scientific knowledge (e) and not to provide a basis for an outcome (g) is enough to avoid responsibility. There is also the practical question as to whether the pure researcher is actually able to foresee the outcome. With respect to the 'philosophical question', I have already suggested that negligent and reckless action can entail responsibility for an unintended effect. More can be said about this matter, but it really separate from the practical question, which is our concern in what follows.

Blame

Following Kant, a number of philosophers have held that persons are praiseworthy only if they act for the *sake* of doing the right thing, and that it is not enough merely to do the right thing. This is surely correct, and it is relevant to our present concerns in that ignorance would seem to preclude any possibility of praiseworthiness, for the following reason. If P does not foresee g as the outcome of her work, and if g is morally significant, then it is still open to us to blame her for her lack of foresight.⁶ There is surely no other way to press the question of moral responsibility. Granted that it is a fact that P is ignorant of g, then either the question is closed or we try to see whether indeed 'ignorance is no excuse'.

This does not work the other way, for surely P cannot be *praised* for not knowing g. This is not to say that it might not be good that P does not foresee g. Suppose it turns out that g is an unqualified good, but that it has no immediate benefit to the ambitious, selfish but nevertheless talented P, and that had P known her work would have this outcome, she would have given up on the project. We judge that it was good that P was ignorant of g. P's ignorance of g being a good thing clearly does *not* imply that P herself is praiseworthy. If g is in some sense evil, whatever this might mean, and P therefore deliberately refrains from coming to know g, then her ignorance is also a good thing: that is, it is good that P failed to see that the consequence of her discovering some item of scientific knowledge (e) would lead to the outcome g.

Decisions not to what is wrong are not, however, themselves worthy of praise.⁷ It is quite another matter if P avoided g at great personal cost, but that leads us back to the issue of freedom and coercion. Finally, to deny all this and say that P is praiseworthy just in case the best, or a superior, outcome eventuates from her actions regardless of what she knows and intends reduces moral responsibility to luck, and that will not do. It is, on the other hand, *prima facie* possible to blame P for 'looking the other way' and failing to try to see what the applications of her work might be.

EXCUSES

Kinds of Excuse

Let us suppose that it is true that P should not have done f because it led to g and overall g was a bad thing. What would excuse P? I suggest that there are three lines of defence, as we might call them, that she can appeal to. All involve an appeal to ignorance on her part, but ignorance of different things. Thus, P might say that she did not know that science has any affects on people and society: she might deny

what was referred to above as a underlying rationale of STS. I will call this, for want of a better term, the 'conceptual defence', as it denies knowledge of what has the status of conceptual connections - between science and society, etc. We will see that this can be dismissed in one short paragraph. Or she might acknowledge this but maintain it is all up to politicians and industrialists. This is the 'political defence'. Then there is what I will call the 'empirical defence'. In this case, P simply denies that she was in a position of foresee g.

The 'Conceptual Defence'.

Let us consider first P's assertion that she is not aware that science affects society. Made in the late twentieth century, this is altogether implausible, for the impact of science on society is manifest to anyone with a rudimentary knowledge of science. In Newton's day, or even in Maxwell's, a scientist might plead innocent in this way, but not now. Again, someone with no formal education might not be aware of the impact of science, but we are talking here of a research scientist. It is a better defence, then, to admit that science affects society but maintain that someone else is to be held responsible.

The 'Political Defence'.

Let's go back to the mythical Beelzebub Industries and the anti-personnel laser weapon. The process of bringing this into existence would presumably require a contract from the military, compliance by the executives of the company and quite possibly the OK of government officials, as well the work of the technical, scientific and production units. We could say that many distinct causal processes have to intersect before the weapon comes into existence. It might be possible to rank these in order of importance in some way. For example, it is possible that many companies are willing to make the weapon and that a particular general's opinion is crucial. It could then be argued that that makes him more responsible than any other individual, but it does not follow that the others, including the working scientists, bear no responsibility at all. However, the excuse canvassed before was, in effect, that only *some* of the causal processes are such as to entail responsibility - those that originate with the military or the politician perhaps.

It is not possible to advance this excuse and at the same time acknowledge responsibility for one's actions when they are sufficient for the outcome in question. In other words, it is not possible to acknowledge sole responsibility, in some instances, while denying co-responsibility in others. Why not? The answer is that one's actions are, in both cases, *necessary* for the outcome in question and we are concerned here with *preventing* wrong. So, from P's perspective, *not* doing f is enough to prevent g, *regardless* of whether or not it would be sufficient by itself to bring about g.⁸

If the whole issue comes down to P denying that she ever has responsibility for anything, then either this is a manifestation of some physically disabling condition or it is not. If P is a psychopath as a result of a brain lesion, then she cannot be held responsible for her actions. Otherwise, P is worthy of blame in a rather fundamental way. An adult who refuses to take responsibility for her actions is not someone that we would admit into our social community: responsibility is a pre-requisite for treatment as a trustworthy equal in any form of social grouping. Thus the blame which attaches to refusal to take on responsibility amounts to exclusion from the group - hence its special quality.

The 'Empirical Defence'

The empirical defence, or 'Polanyi excuse', has it that g could not in fact be foreseen by P . It can be conceded that some highly theoretical research is unlikely to have foreseeable outcomes. So, for example, the pioneers of laser physics cannot be held responsible for anti-personnel laser weapons, Einstein cannot be held responsible for applications of $e = mc^2$, etc. If we make a distinction between normal and revolutionary science, then we could say that the results of the latter are certainly unforeseeable and hence, for such research, ignorance is an excuse. There is, I think, just one way to challenge this view and that is to take a radical Green anti-science line which holds that all science is wrong because it is responsible for the environmental crisis. Science therefore cannot do any good and no one should do science. This is certainly not a line I could defend here.

The empirical defence as a general claim that P can never foresee g is, however, simply false. To show it to be false all that is needed is to find one instance of research leading to an outcome g that could have been foreseen by a little thought and reflection. Of course this is necessarily a little vague: just how much thought and reflection is appropriate cannot be precisely determined. However, if we go back again to the history of the atomic bomb and contrast the attitudes of Szilard, who proposed an embargo on publishing results on fission research in 1938, and Joliot-Curie, who ignored him, we see the difference between a scientist who did think and reflect about the applications of his work and one who did not (at least not in 1938). Thus Polanyi's double unpredictability thesis is simply false: Szilard was indeed able to predict the applications of the work in progress in nuclear physics in 1938.⁹

Granted that the conceptual defence fails, then this in itself is grounds for an obligation on P to engage in a certain (appropriate) amount of reflection and thought about her work. It does so because P cannot consistently acknowledge overall principles about the effects of science on society, about the attendant responsibility of the scientist, etc., and then refuse to act in accordance with them. If, however, the empirical defence worked, then as a matter of fact this obligation would be lifted. But the Polanyi excuse is no excuse, as we have just seen. So, it *is* possible for a pure scientist to be (indirectly) culpable for the ways in which her work affects others because she can be (directly) culpable of being ignorant about these affects. And where there is the possibility of blame and responsibility, there is also obligation. In this way we can argue that scientists do indeed have an obligation to try to foresee the applications of their work and that, in some cases, they bear responsibility for the affects of their work on the rest of us.

CONCLUSION

It is not the purpose this paper to carry the discussion further and seek to formulate instructions for scientists to help them foresee the possible applications of their work. Doubtless this will not be an easy task. The aim here has been to argue that, in general, ignorance is not an excuse when it comes to blame for outcomes that have had bad or detrimental effects on society and individuals. Ignorance is not an excuse precisely because scientists can be blamed for being ignorant.

BIBLIOGRAPHY

Forge, J. (1998) "Responsibility and the Scientist" in Bridgstock *et al. Science, Technology and Society*. Cambridge: Cambridge University Press.

Polanyi, M. (1962) "The Republic of Science" in M.Grene (ed.) *Knowing and Being*. Chicago: Chicago University Press.

Rhodes, R. (1986) *The Making of the Atomic Bomb*. New York: Simon and Schuster.

Ronayne, J. (1983) *Science in Government*. London: Edward Arnold.

Zimmermann, M. (1997) "Moral Responsibility and Ignorance" *Ethics*, 107.

Zimmermann, M. (1988) *An Essay on Moral Responsibility*. Totawa, N.J.: Rowman and Littlefield.

ENDNOTES

1 Polanyi is talking about something being *unpredictable*, literally, cannot be predicted. To take him at his word, we would have to try to see what sort of modality he has in mind. I shall not do this, but take him to mean something like "in the normal course of events, very hard to have predicted".

2 A little more by way of background can be found in Forge 1998 and Zimmermann 1988.

3 This follows Chapter Two of Micheal Zimmermann's *An Essay on Moral Responsibility*. This book, together with Zimmermann's paper "Moral Responsibility and Ignorance", have been most helpful for the present discussion.

4 So the class of actings to which responsibility can attach will *not* be merely a subclass of the class of 'intentional actings'.

5 It is worth raising the issue of recklessness and responsibility again. Suppose I drive fast and dangerously along the highway, in such a way that I significantly raise the chances of an accident. Suppose one occurs. If I just like driving fast, but not causing accidents, then I am directly culpable for reckless driving and (therefore) indirectly culpable for the accident. On the other hand, if I am driving a dying man to hospital, then it would be a moot point as to whether I was culpable for reckless driving. If not, then I would not be responsible for the accident: the 'connection' would have been broken.

6 Blame in this sense does not necessarily amount to judgement by someone in moral authority who *is* aware of the outcome of P's work. In other words, the analysis that we want to give of the conditions under which P can be held to account does not have to incorporate a blamer. Again, just what blame amounts to will depend on one's preferred normative theory.

¹Polanyi is talking about something being *unpredictable*, literally, cannot be predicted. To take him at his word, we would have to try to see what sort of modality he has in mind. I shall not do this, but take him to mean something like "in the normal course of events, very hard to have predicted".

²A little more by way of background can be found in Forge 1998 and Zimmermann 1988.

³This follows Chapter Two of Micheal Zimmermann's *An Essay on Moral Responsibility*. This book, together with Zimmermann's paper "Moral Responsibility and Ignorance", have been most helpful for the present discussion.

⁴So the class of actings to which responsibility can attach will *not* be merely a subclass of the class of 'intentional actings'.

⁵It is worth raising the issue of recklessness and responsibility again. Suppose I drive fast and dangerously along the highway, in such a way that I significantly raise the chances of an accident. Suppose one occurs. If I just like driving fast, but not causing accidents, then I am directly culpable for reckless driving and (therefore) indirectly culpable for the accident. On the other hand, if I am driving a dying man to hospital, then it would be a moot point as to whether I was culpable for reckless driving. If not, then I would not be responsible for the accident: the 'connection' would have been broken.

⁶Blame in this sense does not necessarily amount to judgement by someone in moral authority who *is* aware of the outcome of P's work. In other words, the analysis that we want to give of the conditions under which P can be held to account does not have to incorporate a blamer. Again, just what blame amounts to will depend on one's preferred normative theory.

⁷Some virtue theorists think otherwise, but I assume they are very much in the minority.

⁸ I have not considered the situation in which the may be substitutes for P but not, say, for the general. Consequentialists would typically say that if Q, a 'substitute scientist', would do f, and so bring about g, if P had not, then P is absolved of responsibility because the consequences are the same. I disagree, but fortunately I do not have to address that question here.

⁹The story is well told in Richard Rhodes *The Making of the Atomic Bomb*.