PHIL 530 Keith Douglas
w/ Prof. Paul Russell 13714993

Fischer & Ravizza's Account of Moral Responsibility

Introduction

Fischer and Ravizza (2000 [1998]) defend a compatiblist account of moral responsibility. In this paper I will (section 1) give a brief *précis* of their work, (section 2) expose a genus of possible metaphysical weaknesses in the work, and (section 3) hint how this weakness may be remedied without too much violence to their stated goals and framework.

Section 1 - Précis

This section summarizes Fischer and Ravizza's (2000 [1998]) (hereafter, F&R) In this, F&R defend several aspects of an account of moral responsibility. Page numbers throughout refer to this book unless otherwise stated.

Their account starts by disentangling causal responsibility from moral responsibility (pg. 1-2). The latter allows for various reactive attitudes. F&R's stated aim is to allow us to continue on in our practices of these attitudes and explore when and where they are appropriate. Pages 3-4 stresses the importance that F&R will place later on agents being in control of themselves. Notions of control and of reactive attitudes are regarded by F&R as being central to human moral lives. Hence they are regarded as "given" for their account; any account which failed to accommodate these evaluative practices or also failed to go along with a sort of "moral metaphysics" embodied in their accounts of control, would accordingly be regarded as unsatisfactory by F&R. Chapter one makes clear that F&R's account should be regarded as an articulation of the prereflective notions held in modern, western democratic societies. As such their project is thus to achieve a reflective equilibrium on these issues (pg. 10).

F&R sketch two challenges to these practices. These arguments argue from the assumed truth of causal determinism to the futility of the above practices. The first of these is called the Indirect Argument. This genus of arguments includes several species; one F&R mention on page 18 as involving the *principle of the transfer of powerlessness* (PTP). This involves the idea that the past is out of our control, or is "fixed" in some way. It also embodies another principle, that of the fixity of the laws of nature. Pages 20-21 shows what happens if the laws of nature are exclusively causal. Since humans are powerless over the laws of nature, and the antecedent causes "produce" what

they in fact do, they are powerless to do anything but what they in fact do.

Page 24 introduces a direct argument parallel to the indirect ones above. One is not morally responsible for antecedent conditions distant in the past, and no one is morally responsible for the causal laws that link any arbitrary point in the past to an action in the present. Thus no one is morally responsible for any action.

Chapter two of F&R concerns how agents can be responsible for actions in spite of the above argument. This chapter demonstrates that alternate possibilities are not needed for moral responsibility. A scenario pumps the intuition that despite the fact that someone cannot have done anything but kill the mayor, he is nevertheless morally responsible. F&R acknowledge that this is an usual case as involves what they call a "counterfactual intervener" involving an "alternate sequence". It is unusual because we don't normally believe we are potentially affected by such influences. F&R suggest that two different notions of control are at work, thus the distinction between guidance control and regulative control. An agent possesses guidance control if an action involves free performing of the action. Regulative control is possessed if an agent possesses the power to freely do an action and the power to freely do something else instead. One can thus have guidance control without regulative control. Chapter two also defends the notion that guidance control is compatible with determinism. This focuses around what F&R call "reasons-responsiveness" (pg 37). An agent is reasons-responsive if there are reasons which would "make" the agent do otherwise. F&R point out that there are various subtypes of reasons-responsiveness, and that "Frankfurt-type" examples illustrate the importance of looking at an actual sequence to see if an agent is reasons-responsive. Two dimensions of control, reasons-responsiveness and "mechanisms being agent's own" are independent, and F&R pledge to resolve issues surrounding each separately.

On page 41, F&R introduce an important distinction. This distinction concerns two sorts of reasons-responsiveness, weak and strong. As they say concerning strong reasons-responsiveness:

> "Suppose that a certain kind *K* of mechanism actually issues in an action. Strong reasons-responsiveness obtains under the following conditions: if *K* were to operate and there were sufficient reason to do otherwise, the agent would recognize the sufficient reason to do otherwise and thus choose to do otherwise and do otherwise."

This definition includes three parts. First, an agent must recognize reason(s); second the agent must choose to act, and thirdly the agent must

indeed act. Failures of this “system” can occur at any of these places. F&R regard this strong form of reasons responsiveness to be too strong for guidance control as it opens up an unfortunate possibility. This case is illustrated by a thought experiment involving “Jennifer”. Jennifer decides through an ordinary human process of deliberation to go to a basketball game. Suppose there was a possible reason that would make her weak-willed, but still decide to go. We would want to say that Jennifer has freely chosen to go to the game, is responsible for going, and yet is not strong reasons-responsive. F&R suggest that strong reasons-responsiveness is a too stringent demand for a condition of moral responsibility.

Because of this result, F&R explore weak reasons-responsiveness in the rest of this chapter. An agent is said to be weak reasons-responsive if there is some possible scenario in which an agent has a reason to do otherwise, chooses to do otherwise, and in fact does otherwise. This is illustrated by several examples, including a revision of the aforementioned “Jennifer” case in which Jennifer would decide not to go to the game if the ticket price is $1000 per seat. In this case, Jennifer is said to be weakly reasons-responsive.

To reinforce the utility of the account, F&R apply it to the case of a literally irresistible urge, that of a drug addict. Since there is no scenario involving a reason that would get a severe drug addict to not take his drug, he is not weakly reasons-responsive, and hence not morally responsible for taking that drug. He, however, as they point out, may be morally responsible for becoming an addict. This sort of intuition is elucidated in F&R’s section on tracing, where an exploration of historical issues surrounding moral responsibility is spelled out in outline.

Next, F&R point out that the account sketched in the previous part of the chapter has two features of note. One, it does not require that the entire past history of the world get involved in an agent’s mechanism, thus partially severing the “transfer of powerlessness” worries. Two, “actual sequence” properties of agents are dispositional or modal and thus may require analysis in terms of other possible scenarios.

Chapter three concerns moderate reasons-responsiveness. This is motivated by showing that weak reasons-responsiveness is too lax. Illustrating this is a case of a person with a nervous tic who destroys a valuable artifact. There is presumably a scenario in which the person will respond to a reason and avoid destroying the artifact, this person is thus weak reasons-responsive. F&R say

we have strong intuitions that this person is not responsible for their actions. This result is bolstered by considering “weird pattern” cases. Consider a case involving Jennifer, a basketball fan. Jennifer is going to the game, and is weakly-reasons responsive as there is a reason, say, ticket price of $1000, that would result in her not going to the game. Now suppose one figures out that she would still go to the game if the price were $1001, $1002, and so on. We would like to say that she is not responsible for going, as her pattern of not going is “weird”. This can be extended into more pathological cases. With these in mind, F&R postulate that in order to be morally responsible for an action, an agent must be moderately reasons-responsive. This involves weak reactivity to reasons. Weak reactivity to reasons involves choosing in accord with reasons under at least one possible scenario. Because a choice need not lead to action this is a different requirement than the previous two kinds of reasons-responsiveness.

F&R suggest that this species of reasons-responsiveness involves a regular receptivity to reasons. This involves the agent being required to understand that certain reasons are stronger than others, and accepting a weaker reason entails accepting a stronger one. This also involves having an understandable pattern in this receptivity, as seeable by a “third party.”

F&R test this account with cases involving children, smart nonhuman animals, and psychopaths. These classes of individuals do not contain morally responsible agents as the agents do not have understanding of moral reasons. These individuals may respond to prudential reasons, thus F&R refine their account of responsiveness to require moral responsiveness. Further, the moral responsiveness of an agent should recognize not only that moral obligations apply to others, but also to the agent. F&R sketch boundary cases for their account rather than giving details of what this recognition would involve.

Chapter 4, F&R’s account of responsibility for consequences, begins with a puzzle. Some agents seem to be responsible for the consequences of their actions even if they could not have done otherwise, and some do not. F&R take it upon themselves to disentangle these cases. This is presented using a framework of “consequence particulars” and “consequence universals.” An agent can be responsible for a particular case because she is responsible for the “consequence universal” (e.g. that the mayor is shot). But if an agent is not responsible for the universal in a given case, she is not responsible for the “consequence particular” that may result. In order to understand responsibility, F&R explain how one can be responsible for consequence

universals by using a responsive sequence. This comes in two parts. First, the mechanism that leads to the bodily movement; second is the process leading from the bodily movement to the external world. Both of these must be moderately reasons-responsive. Next F&R present (pg. 120-121) the account of guidance control for consequence universals.

Chapter 5 is F&R's account of responsibility for omissions. The chapter begins with a puzzle over whether being responsible for omitting to do A requires having the ability to do A. Some cases suggest it is required, some suggest the contrary. F&R's attempt to resolve these cases makes use of (pg. 132) the *Symmetric Principle for Moral Responsibility* (SPUR). This principle (broadly) allows negative agency to be treated in the same way as positive agency was previously.

Chapter 6 tackles the direct argument (above). Preemptive over determination cases are used to pump the intuition that although an agent may be vulnerable to transfer of lack of responsibility one wants to say at certain times she is responsible[1].

F&R then (pg. 167) consider an objection to their account, a strong version of the direct argument mentioned earlier: (1) If p obtains and no one is even partially morally responsible for p; and (2) if p obtains, then q obtains, and p leads to q via a causally deterministic sequence; then (3) q obtains, and no one is even partly morally responsible for q. F&R dismiss this as being "straightforwardly unacceptable to anyone who is inclined towards compatibilism."[2]

Chapter 7 concerns how agents "possess" mechanisms. This account relies on responsibility being an historically normative moral phenomenon. These kinds of phenomena work "recursively." If (1) an initial situation is moral, (2) moving from one situation to another "preserves moralness" then (3) all future situations are moral. F&R's account of moral responsibility is genuinely (as opposed to merely epistemologically) historical. Here F&R refer us back to an earlier case when an agent has guidance control over an action that will result in him not having guidance control in the future. The next chapter explains how an agent obtains guidance control. This occurs as part of a

[1] If this sounds implausible or incomplete, it is quite likely that the reader shares Coleman's (2000) worry about this situation. I shall return to this in due course in section 2 of this paper.
[2] As remarked in footnote 1 above, this dismissal seems to be blatantly question begging.

process of moral education. Moral education includes an instilling of ability to respond to reactive attitudes. This response must be appropriate to the evidence associated with the attitudes in question. Pages 225-226 stress that this account of moral education is compatible with causal determinism.

The remaining part of this chapter and the rest of F&R's book can be looked at as refining in small ways the previously developed points in irrelevant ways for the present purpose.

Section 2 - Metaphysical Objections

F&R's book, as can be gleaned from the above précis, is generally more "moral" in tone than "metaphysical". Despite this, I feel that some of the root problems with the account they have sketched are metaphysical. Because of this, and the general principle (which I will not defend) that metaphysics is always epistemically prior in some sense to ethics, I will raise one metaphysical objections in this section of the paper. There are others, but I regard this as the most important one. It concerns the nature of the mechanisms proposed. I move that these are vague and unspecified, and lead to the question begging nature of the response to the strong form of the direct argument on page 167, noted above.

Let's look at in some detail what mechanisms F&R claim are in place in a free-willed agent, namely guidance control. An agent has guidance control if and only if the mechanism that issues in the action is the agent's own, and is moderately reasons-responsive. There are thus two possible genuses of problem here, namely one with the ownership condition and one with the responsiveness condition. I shall consider each in turn.

Chapters 7 and 8 of F&R's book concerns the ownership condition, and as we have seen, they focus primarily on the importance of the origin of action and moral education.

If children are educated into behaving morally, they clearly did not choose to do so - they are not responsible for having a moral education, at least to begin with. And yet we hold people responsible for their actions. This is sort of a transfer of powerlessness argument with regards to moral education. In particular, how is possible that moral education began in the first place. Can something non-moral give moral education? This is intended to "short circuit" the tracing views. Why do we stop the tracing at some point in a child's development, say? Or more bizarrely, even later? If someone becomes a drunk,

how do we decide whether she is morally responsible for having done so? There's no sufficient reason to say "stop the trace here", because there may be influences that affect the eventual consumption in the distant future. Consider something like: "My great grandfather was a little poor, so my grandfather had to work at a backbreaking job as a boy and never got a good education. Hence my father grew up in an even poorer household, and turned to drinking a little too much so in turn I didn't receive enough education about sobriety, so I occasionally drink one too many and drive. In turn then I am not fully responsible for hitting that cat of yours." Why do we consider this implausible? Well, because we think that the influence of one's great grandfather is so minuscule on the situation of hitting a cat while slightly drunk in a car. But without explaining why moral education (of the "good sort"[3]) stops this. Note carefully this account does not rely on complete lack of moral education at any point. Note also I am not saying that a "just so story" like the above would ever mitigate all moral responsibility[4] .

The above is similar to the worries of "ultimate responsibility" raised in chapter 5 of Kane (1996). Of course the agent's mechanism in the above account is in some sense hers - it is in her body, after all. But the agent wants to be partially responsible for creating it, and not having it created at the mercy of the environment or other agents.

A similar issue applies to the responsiveness condition; this time the issue concerns individuation and responsiveness of actions. F&R suggest that for an agent to be morally responsible requires that their preference structure obey some sort of partial ordering. However, this involves several key metaphysical that perhaps preference structure of agents in the actual world do not obey completely. Most people would acknowledge that <u>in a particular domain</u>, that stronger reasons in the sense of more costly reasons (e.g. $1001 ticket to a basketball game is a stronger reason for not going to the game than a $1000 ticket, etc.) outweigh weaker reasons, etc. But what of "incommeasurable" reasons across domains, and further, how does one rank reasons at all in certain domains?

On the former point: how does one compare (e.g.) the value of attending a

[3] Part of the concern here is what the details of what F&R take to be good moral education are very sketchy.

[4] This points to a general problem with F&R's account that I do not have time to go into. It strikes me that moral responsibility is not a unary predicate, but of at least the fourth degree (moral responsibility to person x in respect y to degree d at time t), assuming degrees of moral responsibility could be given at least ordinal ranking.

basketball game vs. helping a friend to study for an exam. There are cases where our reasons-responsiveness would be very clear, but introspectively (and from my discussion with others) it seems that boundaries between preferences become "chaotic" - that small differences in situations lead to wildly different preference rankings. Consider the case of someone working in their office at 5:55 on Friday. Finishing a report now ranks higher than going home to eat dinner. But a short time later, at 6:00, the rankings are reversed. We can understand this pattern because we are used to people being very literal about "quitting time", and further I have even provided the times in this case. But in more complex cases, especially on incomplete information, there is a great danger of misidentifying a reasons-responsive mechanism as the mechanism's behaviour may be chaotic in the mathematical sense (i.e. very sensitive to small changes in initial conditions, etc.).

On the latter point: the Arrow paradoxes[5] (while perhaps misapplied often in social studies (Bunge 1998)) seems to possibly apply to many individuals at the level of individual preferences. There is both the danger here of having differing states of minds leading to different ranking, and thus leading to the problem of an outsider determining which evaluation scheme was used in any case, and hence a problem in evaluating responsiveness. But there is also a danger of an infinite regress on the part of the agent. Because now, her methods of evaluation can themselves be subject to evaluation, and, perhaps, even the choice of mechanisms she is going to use (if we assume that there are in some sense several mechanisms that can lead to action). F&R have not specified much about the mechanisms that would help decide these issues. This objection is not meant to be "knock-down", but just simply suggest a theme for further work, of which I will sketch one route in the next section.

Section 3 - Remedying the Weaknesses

The fact that some of the weaknesses in F&R are metaphysical does not entail that their solution is such. Here I am going to suggest a pragmatico-legal solution[6], rather than a metaphysical one.

In the case of the responsiveness condition, I suggest that responsiveness be regarded extremely falibilistically. If moderate reasons-responsive mechanisms

[5] Consider three 'states of mind' (i.e. ways of ranking preferences) of an individual: 1,2,3. Suppose that these states of mind rank choices A,B,C. Then let 1 produce A>B>C, 2 produce B>C>A and 3 produce C>A>B. Then collectively A>B>C>A, unfortunately.

[6] I am doing so in the interest of charity and time. If the present author had more time to work on the issues more, he would possibly be willing to state that F&R's book needs its entire metaphysics gutted.

are likely to be chaotic, as I have suggested, it stands to reason that until we know more about them we should treat them as being potentially random (as chaos in the technical sense is known jokingly as “mock randomness”). This means we have to be prepared to be mistaken relatively frequently about whether agents are moderately reasons responsive. Thus my suggestion here is to err on the side of caution, and give people the benefit of the doubt, until we know something more about the mechanisms. (This does fit in, I think, with our general practices about “innocent until proven guilty”, and so on.

The “ownership” issue is a bit more difficult. In this case I suggest that responsibility be explicitly taken by individuals at a certain age or level of psychological development (preferably the latter, but that requires developments in social psychology and so on that are not immediately forthcoming.). We basically allow someone to choose whether they accept broadly speaking the notion of being morally responsible. Thus we can stop the tracing at some socially sanctioned point in the past. Perhaps one should build into this a test of the success of the moral education. (This test’s results of course would be revisable in the light of future information or changes in the agent.)

This latter suggestion is not well developed at all, but I think it is one that is important in any account of taking responsibility regardless of the degree to which one takes F&R’s account seriously.

Conclusion

I have sketched F&R’s account, raised a family of objections, and suggested possible directions to take the account in order to deal with these. Further work would include following up on these directions and seeing if the largely implicit metaphysics of F&R’s would hold up these (and, no doubt, other) tasks.

References

Coleman, K. 2000. Unpublished remark, proceedings of University of British Columbia Course PHIL 530, Winter 1999-2000 Session.

Fischer, J. M. & Ravizza, M. 2000. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Kane, R. 1996. *The significance of free will*. New York : Oxford University Press.