Consequentialism, Strawsonianism and the illusion of undermined responsibility

Gunnar Björnsson
Linköping University
University of Gothenburg

Scepticism and consequentialist about moral responsibility

Two features about our everyday practice of holding people responsible seem to tug in opposite directions. The first feature is that our attributions of moral responsibility for decisions, actions and outcomes as well as our practice of holding agents responsible are notoriously sensitive to sceptical arguments. Ordinarily, people take agents to be morally responsible for their actions and take them to deserve blame or sanctions for bad actions or praise and rewards for good deeds, without prior reflection on possible metaphysical prerequisites for moral responsibility. But they often come to see metaphysical considerations as highly relevant and find their confidence in moral responsibility shaken when introduced to regress arguments such as Galen Strawson’s “basic argument” or Peter van Inwagen’s “direct argument”, arguments from manipulation such as Derk Pereboom’s “four case argument”, or arguments from luck, such as Al Mele’s contrastive argument.

The second feature is that our practice of holding people responsible is largely driven by concerns about how motivational structures are affected by our holding or not holding people responsible for decisions and their outcomes. The most obvious sign of this concern is that people often motivate practices of holding people responsible with reference to what would happen in their absence: people would care less about values beyond their immediate interests, go lazy, engage in free-riding. Somewhat more subtly, it is clear that our reactive attitudes, expressed in our ways of holding agents responsible, are sensitive to the qualities of will of those agents. If we learn that an action was not the result of ill will, our tendency to hold the agent responsible for a bad outcome tends to be diminished, just as one would expect if the concern were to modify faulty motivational structures. Expressions of regret and guilt and thus willingness to change motivational and behavioural patterns tend to placate indignation or resentment. From an etiological perspective, it seems plausible that our species have reactive attitudes and engage in practices of holding each other responsible exactly because such reactions modify motivational structures and behaviours in ways that protect and promote values that we care about.

To the extent that our holding people responsible is motivated by the effects of holding people responsible, it is puzzling why it should be subject the sceptical concerns: neither of the sceptical arguments mentioned above seem to undermine the usefulness of holding people responsible. This is how the two features seem to tug in opposite directions.

In his influential paper “Freedom and Resentment”, Peter Strawson argued that we should let our practice of holding people responsible be deeply affected by neither of these
concerns. Unlike “pessimists”, we should not be moved by sceptical, incompatibilist, concerns because they involved judging the practice from a metaphysical perspective foreign to the participatory stance to which our practices of holding people responsible belongs. Unlike “optimists”, we should not justify our practice with reference to its effects, because, again, such concerns are external to the practice itself, or at least leaves out concerns that are internal to the practice, concerns that are focused on how the action came but have no truck with consequences of holding the agent responsible—perhaps the agent is dead.

In my paper, I argue that Strawson and some of his followers misrepresent the relation between sceptical arguments and our practice of holding responsible. As many of his critics have pointed out, the appeal of such arguments is very natural and almost unavoidable given the shape of the concept of moral responsibility that governs this practice. Unlike many of his critics, however, I will argue that our impression that the considerations invoked in such arguments diminish responsibility is nevertheless an illusion, comparable to other cognitive and perceptual illusions. Similarly, I will argue that Strawson misrepresents the role of “external” consequentialist concerns. It is correct that our attributions of moral responsibility are backward-looking, relying on information on how a decision, action or outcomes came about rather than on potential effects of holding the agent responsible for it. But it is also true that our judgments focus on backward-looking concerns because our practices of holding responsible is largely driven by forward-looking—consequentialist—concerns.

The argument proceeds in three steps. The first is to make plausible an empirical theory about the concept of moral responsibility operative in our practices of holding people responsible and attributing desert. The second step is to explain why sceptical arguments have intuitive force given this structure. The third step is to present a plausible account of why a concept satisfying this empirical description is governing our judgments of moral responsibility as well as our practices of holding people responsible. The fourth step, finally, is to argue that intuitions resulting from sceptical arguments are best understood as illusory because they are insensitive to what our concept of moral responsibility has been designed to track.

The Explanation Hypothesis

Elsewhere, I have introduced and defended an empirical hypothesis about how we make judgments of moral responsibility. Simplifying things, it says that:

*The Explanation Hypothesis*: People take agents to be morally responsible for something insofar as they take it to be straightforwardly and saliently explained by a motivational structure of the agent that is of a type that tends to be affected by expressions of reactive attitudes in the right way.

The motivational structure *saliently* explains an event (a decision, action or outcome) if its role in a full explanation of that event is particularly central given our explanatory interests. Everyday explanatory judgments are selective and pick out conditions in the causal prehistory of the explanandum that are especially relevant given our prior expectations (we do not cite
the presence of oxygen in the air as explaining why a house burned down), given our contrastive interests (we do not cite smoking when explaining why one of two heavy smokers got lung cancer but not the other), and given our normative expectations (if two people jointly cause an accident by behaving in the same way, the person who \textit{should not} have behaved in that way is seen as “the cause” of the accident).

The selective nature of everyday explanations makes the Explanation Hypothesis a surprisingly powerful account of judgments of moral responsibility. Obviously, the hypothesis can account for the fact that we take people to be responsible for most intended outcomes of their actions: because of our powerful goal-directed mechanisms, such outcomes are straightforwardly explained with reference to what we want to achieve. But relying on the selective nature of significant explanations, it also provides a unifying account of how a wide variety of otherwise disparate phenomena affect judgments of responsibility. As I have argued elsewhere, it explains why we take it that (a) external force, (b) threats and (c) ignorance mitigate moral responsibility to various degrees, as well as why we take it that (d) those who actively participate in the production of an outcome have a higher degree of responsibility for it than those who merely allow others do it, that (e) someone who takes initiative is more responsible than someone who tags along, and that (f) agents are more responsible for known negative than for known positive side effects that the agent does not care about (Björnsson and Persson 2009; forthcoming). It also makes sense of cases where several agents are \textit{jointly} responsible for outcomes that neither of them could have prevented (Björnsson, forthcoming) and explains a variety of results of recent experimental studies of moral responsibility (Björnsson and Persson ms). This explanatory success provides considerable support for the Explanation Hypothesis.

\textbf{Explaining sceptical intuitions}

The Explanation Hypothesis also explains why judgments of responsibility tend to be undermined by considerations suggesting that (g) our decisions are a matter of luck, that (h) our actions are, ultimately, the upshots of events over which we have no control, or that (i) our behaviour can be given reductionistic, mechanistic explanations. By changing explanatory interests and salient explanatory categories, sceptical arguments invoking such considerations undermine ordinary judgments of moral responsibility (Björnsson and Persson forthcoming). What is important here is that the mechanisms that produce these sceptical intuitions are the very same mechanisms that account for various prosaic variations in judgments of moral responsibility. \textit{Pace} Strawson, scepticism based on metaphysical considerations flows naturally from our everyday ways of thinking about of moral responsibility.

\textbf{Conceptual aetiology}

It seems undeniable that the central role of our judgments of moral responsibility is to guide our practices of \textit{holding} people responsible. As mentioned, these practices seem to be largely concerned with affecting motivational structures so as to promote or prevent certain kinds of
events. Assuming that this function has shaped our practices, this provides an eminently plausible explanation of why our concept of moral responsibility has the shape it has. In order for our practices of holding people responsible for various events to reliably perform this function, they need to be directed towards exactly the sort of motivational structures that our concept of moral responsibility tracks under normal circumstances: structures that explain these events in systematic ways, respond in the appropriate way to the agent’s being held responsible and saliently explain the event in question on the occasion such that the connection between motivation and event is clear enough to support learning. This suggests that the concept is designed for its capacity to distinguish between cases where these relations do and do not hold.

The illusion of undermining responsibility

In the central section of the paper, I argue that the Explanation Hypothesis and corresponding conceptual aetiology gives us reason to think that whereas sceptical intuitions flow naturally from our concept of moral responsibility, what they tell us is in fact illusionary, just as the perceptual appearance that an oar is bent when it sticks into the water at an angle is illusionary. Our concept of moral responsibility has the shape it has and plays the role it plays because it helps us distinguish between cases where the relation between types outcomes and the relevant sorts of motivation is and is not salient from stable everyday perspectives. It is when reactive attitudes are targeted at cases where the relation is stably salient that they are likely to reinforce or reform motivational structures in the relevant way. The function of negative responsibility judgments—judgments to the effect that someone is not morally responsible for something—is to prevent us from holding people responsible when the relation is missing. Relative to this function, intuitions triggered by sceptical arguments are misleading: although the relevant relation is not salient given the explanatory interests and explanatory categories that are triggered by sceptical arguments, they might well be from less esoteric perspectives. For this reason, our intuitions of moral responsibility need to be corrected in judgment, just as the appearance that the oar is bent.

In the final section, the argument is defended against a number of objections. The upshot is a view that preserves some central aspects of the Strawsonian approach: sceptical arguments against moral responsibility can be rejected once we have a clear view of the role of judgments of moral responsibility in governing our practice of expressing reactive attitudes and holding people responsible.

Björnsson, Gunnar and Persson, Karl, ms: “Explaining Judgments of Moral Responsibility”, manuscript
Björnsson, Gunnar, forthcoming: “Joint responsibility without individual control—the Explanation Hypothesis”, forthcoming in Compatibilist Responsibility: beyond free will and determinism, eds. Jeroen van den Hoven , Ibo van de Poel and Nicole Vincent