

# *The Emotivistic Concept of Moral Responsibility. A Critical Analysis of C. L. Stevenson's Account\**

Mikołaj Golembiowski

Warsaw University  
m.golembiowski@uw.edu.pl

## Abstract

We will discuss an interesting account of moral responsibility presented by Charles L. Stevenson in his *Ethics and Language*. We will take a closer look at the most important statements and weigh the arguments of the opponents. We will try to answer the question whether Stevenson's account is plausible or if we really need, in explaining our concept of moral responsibility, to refer to some other accounts, especially to the libertarian ones. The problem that will be considered is very closely connected to the compatibilism-incompatibilism debate. If Stevenson's main statements turned out to be implausible, this fact would deliver an argument supporting incompatibilism. Otherwise, we could claim to have found a good compatibilist framework account of moral responsibility.

## 1. Introduction

Moral responsibility is one of the most fundamental problems that accompanies our normative deliberations. Considering if someone is responsible for his action, we often ask if the action was avoidable for that person or if that person could have done otherwise than he actually did. When we come up with the answer that the action he actually did was unavoidable for him we are often prepared to abandon our moral evaluations of such an action. Every reliable theory

\*This paper was previously presented at the 1<sup>st</sup> Philosophy Graduate Conference at the Central European University, Budapest, March 19, 2006. I would like to thank András Szigeti for interesting comments.

of moral responsibility should help us to answer the questions: Why do people tend to treat the actions they believe to be unavoidable in such a way that they abandon evaluating these actions in moral terms? Why do people tend to accept excuses like “I couldn’t help it”, “It was unavoidable”, etc.?

In the following paragraphs we will discuss an interesting account of moral responsibility presented by Charles L. Stevenson in his important work *Ethics and Language*, in the chapter: “Avoidability, indeterminism.”<sup>1</sup> We will take a closer look at the most important statements and weigh the arguments of the opponents. We will try to answer the question whether Stevenson’s account is plausible or if we really need, in explaining our concept of moral responsibility, to refer to some other accounts, especially to the libertarian ones. The problem that will be considered is very closely connected to the compatibilism-incompatibilism debate. If Stevenson’s main statements turned out to be implausible, this fact would deliver an argument supporting incompatibilism. Otherwise, we could claim to have found a good compatibilist framework account of moral responsibility.

Let us recall some fundamental components of Stevenson’s understanding of metaethics, which unfortunately tend to be forgotten by many of his opponents. Emotivism is a theory which is formulated to explain the phenomena connected with our moral language. It does not aim to promote any particular normative claims or to evaluate our factual moral beliefs. This means that it shall not be the aim of the dissertation to make any normative claims about what kinds of action should be excused or in what kinds of situation should we treat an agent as non-responsible. Its aim will be rather to explain how people ascribe moral responsibility to agents and why they do so in the established way. Thus, we will be treating responsibility as a normative concept and explaining it and its natural story in a descriptive, not a normative way. Because in this context Stevenson used to talk mostly about *making moral judgments, judging someone’s action*, etc., and not about *ascribing responsibility*, we’ll treat the latter as equivalent to *being prepared to judge someone’s action in moral terms*.

## 2. Avoidability

Stevenson begins his investigations of moral responsibility by explaining the meaning of the phrase “avoidable action.” He suggests the following definition:

- (D1) “A’s action was avoidable” has the meaning of “If A had made a certain choice, which in fact he did not make, his action would not have occurred.”<sup>2</sup>

The examples of real situations in which people apply their moral judgments, should lend credence to this definition. The author considers the dilemma of a commander who is to judge an army officer that has failed to win an engagement. The commander must decide whether or not to censure the officer. If he were

<sup>1</sup>See Stevenson (1944), 298–318.

<sup>2</sup> See Stevenson *op. cit.*, 298.

convinced that the failure was avoidable, he would rather censure or punish the officer. The given definition helps us explain why this is the case. If the officer's action were avoidable - meaning: the officer's making a different choice could have prevented the failure - then, granted the uniformity of nature, failure will be prevented in any similar future cases if only the officer makes the right decision in those cases (meaning: the same decision that would have prevented the failure the commander is now judging). The commander's censure will probably motivate the officer to make such choices that will allow him to win future engagements. But if the failure was unavoidable in the sense of definition (D1), the commander's censure will probably not help the officer to win in the future.<sup>3</sup>

In his article 'Ethical judgments and avoidability' (Stevenson, 1938), written before *Ethics and Language*, Stevenson proposed a somewhat different version of the definition (D1). He considers a possible counterexample: It could be the case that the commander is convinced that it would not have been enough for the officer to make another choice to prevent the failure, but he denies that the failure was unavoidable. He could think that, for example, the officer would not have failed if he had a sufficiently intense interest in bringing about what he chose. So, Stevenson adds a new condition to the initial definition:

(D2) "A's action was avoidable" has the same meaning as: "If A had chosen a certain different alternative, and if he had had a sufficiently intense interest in bringing about what he chose, then his action would have been prevented."<sup>4</sup>

An essential feature of (D1) and (D2) is that these definitions let us distinguish between avoidable actions and actions that sprang from an indetermined choice. This is because avoidability is concerned with what would have happened if some choice had been made that in fact had not been. It does not imply that the actual choice was not wholly determined by preceding events.<sup>5</sup> This means that Stevenson's account is a compatibilist one. On his view, there is no contradiction between the claim that determinism is true and ascribing moral responsibility to agents. Moreover, the author suggests that some moral philosophers are confusing the question of avoidability with the question of causal determination, and that this initial confusion results in considering a pseudo-problem of no consequence to moral philosophy.<sup>6</sup>

On Stevenson's view, if someone's action fulfils the condition given in (D1), and probably also the second condition given in (D2), we can say about this person that the man is "at liberty" in the sense propounded by Hobbes, or, in Hume's words, he is not acting under constraint.<sup>7</sup> This notion of liberty, popular among compatibilists, assumes that a person is free if he can do what he is

<sup>3</sup> Cf. *ibid.*, 303-4

<sup>4</sup> See Stevenson (1938), 53-54.

<sup>5</sup> See Stevenson (1944), 299-300.

<sup>6</sup> Cf. *ibid.*, 299

<sup>7</sup> Cf. *ibid.*, 229

actually willing to do. This means that the author wants to see his account of avoidability as corresponding to the Humean conception of liberty. The statement “*A*’s action *X* was avoidable” should entail the statement “*A* did *X* under no constraints” and “*A* did *X* under some constraint” should entail “The action *X* was unavoidable for *A*.”

But is this really the case? Is every action performed under some constraint an unavoidable action in the sense of Stevenson’s definitions? Let us consider the conflict between someone’s first-order and higher-order desires. Someone stole something from a shop. The thing he had stolen does not really matter to him. He did it because he is a hopeless kleptomaniac. He had a strong first-order desire to steal it and his second order desire to be a decent person who does not break the law was too weak to prevent his action<sup>8</sup>. Consider now whether to that person stealing was an avoidable action in the sense of definition (D2)? Real problems emerge when attempting to give an adequate explanation of the phrase “sufficiently intense interest.” But if we follow the author’s own suggestion, i.e. that the word “interest” means the same as “any kind of desire, aversion, etc.”,<sup>9</sup> we can with no problem classify the kleptomaniac’s action as avoidable. It seems obvious that if our kleptomaniac had had a sufficiently strong desire not to steal and had chosen to act on that desire, he would not have done what he did. The problem is that his second-order desire was insufficiently intense for him to act on it.

Is the kleptomaniac acting under constraint? Stevenson claims that if an act is avoidable in the sense of his definition, then its performer is not acting under constraint. But if we admit this, then we will have to allow that some compulsive actions do not qualify as constraints. Hume’s own definition of liberty as “a power of acting or not acting according to the determination of will”<sup>10</sup> allows us to count compulsive actions both as actions that are and are not done under constraint, because it is not specified what is meant here by the word “will.” Thus, we can easily interpret following some first-order desire as acting in accordance with the determination of will. But this seems implausible because we think there is something that the kleptomaniac could not accomplish, something which is so important for him that we might call it: “the thing he really wants” – it is his second-order desire to be a decent person - and we can juxtapose this with his first order desire to steal, which is something he cannot resist. We could then reject Stevenson’s claim about the correspondence between his account of avoidability and the concept of constraint. But the real problem is that many people considering similar situations would rather say that both obtain: the agent acted under constraint and his action was unavoidable to him. This suggests that the definition (D2) could be inappropriate.

<sup>8</sup> The concept of first-order and second-order desires was presented by Harry Frankfurt, see his (1986), 65–80

<sup>9</sup> See Stevenson (1938), 54.

<sup>10</sup> See Hume (2004), 61.

There are some other reasons to suspect something essentially wrong with Stevenson's account of avoidability. C. A. Campbell, criticizing this conception and a similar one presented by G. E. Moore,<sup>11</sup> claims that there is a huge set of actions which tend to be regarded as possible to prevent, if only an agent had made another choice, but the question of avoidability of these actions remains relevant.<sup>12</sup> He gives an example of a liar. In most cases it is obvious that a liar would not have lied if he had chosen not to. But it remains an open question whether his action was avoidable for him, or if he could have told the truth. This is because when we are making moral judgments about lying, the thing we are actually interested in may be whether the agent could choose not to have lied. Campbell's claim, understood as a statement about the common usage of word "avoidable", seems to be more plausible than Stevenson's definition. And the point is that we have already declared wanting to treat Stevenson's account as explaining our actual practice of ascribing moral responsibility to agents, and not as moral justification for doing what we actually do. This provides a good reason to admit that Stevenson's concept is inappropriate.

But we should not forget that Stevenson's explanation of avoidability is a part of his broader theory which concerns morally evaluating actions. We can still consider this other part, without making any claims about the correctness of his concept of "avoidability." We will then return to the problems we have just discussed. This should be very helpful in making a reliable assessment as to why these problems with the definition of avoidability occur, and in estimating the losses that a compatibilist theory would suffer if we had rejected this concept.

### 3. The forward-looking function of moral judgments

Let us remember that Stevenson's main purpose was not to provide a theory of meaning for the word "avoidable", but to explain the phenomenon that people restrict their moral judgments to actions they consider avoidable from the agent's perspective. In his opinion, this is closely connected to the fact that moral judgments are not descriptive, but rather express the speaker's approval or disapproval of the action that is judged and induce the hearer to share this attitude.<sup>13</sup> When we are expressing our disapproval of an action, what we really want is that our hearers do not act similarly in the future. Moral judgments are quasi-imperative. Their main function is control and modification of someone's future actions and attitudes. In this sense they are forward-looking. In spite of the fact that they usually figure in the past tense, they actually concern some events that are expected to occur in the future.<sup>14</sup>

To illustrate the forward-looking function of moral judgments we should turn back to the example we have already considered. The commander confronts

<sup>11</sup> See Moore (1912).

<sup>12</sup> See Campbell (1951).

<sup>13</sup> See Stevenson (1944), 21–2.

<sup>14</sup> Cf. Stevenson (1944), 301–2, 306–7.

a dilemma if he should censure one of his officers, who has failed to win an engagement. He might choose to censure the officer to induce him to modify his choice in any future cases, and so the commander's decision will be explained by his present will to win in future engagements. Hence, the decision to ascribe to the officer responsibility for his action is forward-looking.

But let us consider some more complicated examples given by Stevenson. The commander might have some other motives to censure the officer. He may want to give a warning of a general precedent, to prevent similar failures in any of his other officers. Or he may want to cause the officer to resign his commission. In the second case the commander's judgment will not induce the officer to act differently in similar circumstances, but it will rather keep him from encountering these circumstances.<sup>15</sup> This kind of motivation for evaluating someone's actions in moral terms does not require any great changes to Stevenson's main idea. In both cases ascribing moral value remains forward-looking and there is still a close connection between Stevenson's concept of avoidability and the practice of applying moral judgments. If the commander wanted to censure someone for an action that would have been unavoidable to *anyone* in given circumstances, his judgment would be ineffective, it would not prevent any future failures.<sup>16</sup>

Let us shortly sum up the results of our investigations to date. The theory we are examining aims to explain our practice of ascribing moral responsibility to agents by showing that evaluating actions in moral terms is an effective device for controlling one's future behaviour. We can make the following two statements:

- (S1) It is a fact that people do not tend to evaluate actions in moral terms if they do not expect these evaluations to have any bearing on the performance of similar actions in the future.
- (S2) Unavoidability of an action *X* (in the meaning proposed by the proposed definition) indicates that a moral evaluation of *X* will be ineffective in controlling future actions.<sup>17</sup>

As we have already seen, it is not always the case that in judging one's action we want to control the future actions of that person and that person *alone*. Obviously, this is not the case when our moral judgment concerns a conduct of fictional characters or of the deceased. But this fact can be explained with relative ease. When we judge the conduct of literary characters or historical figures we often want to influence the attitudes and future actions of the hearers. For example, if our judgment is negative we are trying to prevent the hearer from taking the character as a model for his own conduct. Or we are strengthening our own resolve not to act like the person we are judging in similar circumstances. Sometimes we may evaluate the conduct of all the persons the given character

<sup>15</sup> Cf. Stevenson (1944), 306.

<sup>16</sup> As above.

<sup>17</sup> Cf. Stevenson 1944, 303.

typifies.<sup>18</sup> In all of the above cases, our judgment remains forward-looking and avoidability of the action (in the sense of Stevenson's definition) remains a condition of effectiveness of the judgment in controlling future conduct of the hearers, of ourselves or of the people the character typifies.

It is worthwhile to emphasize that on Stevenson's view unavailability of an action only *indicates* the ineffectiveness of moral judgments. It is not the case that the statement: "A's action was avoidable", as Stevenson has defined it, implies one's belief that it would be sensible or rational to judge A's conduct in moral terms. This is because there are some situations where if *A* had made a certain choice, which in fact he did not make, his action would not have occurred, yet we think there are only bleak prospects to induce A's conduct by making moral judgments. For example, we often abandon judging some actions of the elderly in moral terms, believing that it will not help to prevent similar actions in the future, because it is too difficult to in any way influence the customs of such persons. (The other reason could be that we believe the person has only little time to act on these customs and it would be very difficult for us and for her to change them, so we decide to tolerate her conduct.) We can, then, formulate the next statement of our theory:

- (S3) Increasing difficulty of controlling conduct *X* by making moral judgments diminishes people's willingness to evaluate *X* in moral terms.<sup>19</sup>

There are also some good reasons to reject Stevenson's unclear and misleading suggestion that avoidability of an action (in the defined sense) is a necessary condition for the effectiveness of evaluating it in moral terms.<sup>20</sup> It therefore seems implausible that unavailability of an action *X* entails the *impossibility* of controlling future conduct by making moral judgment about *X*. First of all, we cannot preclude that someone's moral judgment of an unavoidable action would contingently prevent the occurrence of similar actions in the future. Secondly, judging an action that in the given circumstances *C* would be unavoidable for all possible agents could sometimes prevent an agent *A* from making a similar choice in circumstances *D*, very similar to circumstances *C*, but differing in that the action is avoidable for *A*. We can't forget that agents' ability to recognize some very nuanced differences between circumstances are limited. And all these limits can also be used to control one's future conduct and modify one's attitudes. But we should not let this trouble us exceedingly. All what we need is for the theory to provide an explanation of our actual practices as they are. The fact that the *possibility* just presented is permitted by the theory does not imply any *moral justification* of such practices.

<sup>18</sup> Cf. Stevenson (1944), 302.

<sup>19</sup> Cf. Stevenson (1944), 311.

<sup>20</sup> See as above. There are some a little bit misleading formulations like "We are for the most part wholly unwilling to judge unavoidable actions, seeing that is impossible to control them."

#### 4. Stevenson's account and the compatibilism-incompatibilism debate

Let us return to our principal problem. We can now examine how the theory just considered copes with the problems mentioned in the earlier part of this dissertation. First of all, we may emphasize that the theory is to a large extent independent of the fact that Stevenson's definition is implausible as a claim about the common usage of word "avoidable." This is because the fact that people sometimes abandon ascribing moral responsibility to agents and accept excuses like "It was unavoidable" is explained by their unwillingness to do something they believe to be ineffective. We can treat the expression "unavoidability", as it appears in the statement (S2), as a purely technical notion and replace it with its definitional equivalent. But we must then admit that our definition does not explain the real meaning of the word "avoidable" that appears in the actual excuse we are considering.

Should we really let this trouble us so much? Should we reject Stevenson's forward-looking account of moral judgments as an explanation of ascribing moral responsibility? Not necessarily. We can treat Stevenson's technical notion of avoidability as providing part of the explanation of *why we actually accept* some excuses that contain the word "unavoidable."

Let us again consider the example given by C. A. Campbell. We admit that there is nothing peculiar in asking whether a liar's action was avoidable for him. But is it the case that any answer which aims to show the liar's action as unavoidable for him, in some sense of the word, will constitute a universally accepted reason to abandon evaluating this action in moral terms? This is open to doubt. We could rather expect it would be only very rarely that someone was prepared to accept the following excuse: "This lie was unavoidable for me because it is written in my genes that I lie in circumstances such as those that have just occurred." And there is a high probability that the person who does accept the above excuse has been influenced by some philosophical position.

Stevenson's theory seems to imply that if such an excuse for some action *X* is universally accepted, then this causes the emergence in the hearer of a belief that judging *X* would have no effect on the agent's future conduct. This also concerns the case of a liar. A liar could expect to very probably be excused if, for example, he or someone else were to provide evidence that his action was a result of some incurable cerebral disorder, and that it is practically impossible to induce him to stop lying by censuring his conduct or by threatening to punish him. The same concerns any instances of compulsion, weakness of will, etc. If, for example, a kleptomaniac's conduct was thought to be resistant to every kind of censure, threat and punishment, we could expect that it would be universally accepted to abandon ascribing him moral responsibility. But it seems rather to be the case that sometimes we do treat such persons as responsible agents – we may demand of them, for example, that they undergo medical or psychological treatment or that they avoid situations in which their affliction usually manifests

itself. Censuring them for some of the thefts can sometimes help to make these persons do something in order to prevent certain situations in the future. These facts correspond well with Stevenson's account because they confirm his claim that moral judgments are forward-looking.

And what about libertarian accounts of free will, such as Campbell's? First of all, a libertarian-incompatibilist should give us a credible example of a situation whereby *A*'s excuse for some action *X* will be universally accepted in spite of the fact that people who accept it believe censuring *A* for doing *X* to be very effective in modifying his attitudes and controlling his conduct. Were the incompatibilist to find such an example, which seems a daunting task, some evidence should still be provided to support the claim that the reason why people abandon ascribing moral responsibility to agents in such cases is connected with their belief that this action sprang from undetermined choice. That is, an incompatibilist is obligated to show how an agent, who is not mentally ill and is not suffering from any noticeable personality disorders, could prove or give some evidence that her conduct, for example lying, is determined, as opposed to resultant from free choice. In most cases what incompatibilists actually do is claim that they possess intuitional access to the meaning of phrases such as "avoidable", "could have done otherwise", etc. They say their intuition also tells them that moral responsibility is incompatible with determinism.<sup>21</sup> But what reason do we have to trust their intuitions?

The difference between Stevenson's compatibilist account of responsibility, on the one hand, and Campbell's or Van Inwagen's accounts, on the other, is as significant as that between a scientific theory and some claims of a mystic. The first can be empirically tested, modified and, if it failed to give satisfactory results, easily rejected as degenerating. But none of this hold for the second, which is speculative and makes few (if any) claims that could be empirically tested. This means that rejecting Stevenson's compatibilism in favor of one of the mentioned incompatibilist theories is in fact not acting in compliance with the standards of rational methodology of scientific research.<sup>22</sup>

Stevenson's account should probably be modified in some respects. We could, for example, question whether what makes us unwilling to make moral judgments about some actions are, as Stevenson suggests, our actual conscious beliefs. Instead, we can think, for instance, about some evolutionary adaptations. It is a fact that people do not tend to evaluate actions in moral terms if evaluations of that kind of human conduct do not usually have any bearing on the performance of similar actions in the future. But it is not the case that people always deliberately calculate the utility of making such evaluations or that their dispositions to make ascriptions of responsibility are so closely connected with their actual beliefs concerning that kind of utility. All these dispositions are rather, as Peter

<sup>21</sup> See for example [Van Inwagen \(1986\)](#).

<sup>22</sup> The methodology was formulated by Imre Lakatos, see his (1978). I'd like to thank Adrian Kuźniar for bringing it to my attention in this context.

Strawson maintained part of the general framework of human life.<sup>23</sup> Strawson seems to be right claiming that in moral judgments and ascriptions of responsibility people express interpersonal attitudes which are usually initial and primitive and are not supported by any sophisticated reasoning. Let us notice, however, these attitudes and dispositions are part of human nature which evolved. They were confronted with attitudes of other human or pre-human beings and external environmental factors. We are descendants of survivors. We inherited dispositions which turned out optimal for social cooperation. It seems not so difficult to explain why our attitudes tend to fulfill the conditions presented in the statement (S3). It was simply evolutionary profitable to resent, blame or praise, punish or reward others only if it influences the conduct of the members of one's community. We may admit Strawson that the practice of ascribing moral responsibility as a whole does not call for an external "rational" justification<sup>24</sup>, but we should also notice that rationalisation presented by Stevenson and other simple-compatibilists, like Hume, Schlick, Ayer, helps explaining the evolutionary genesis of some of our moral practices.

Maybe Stevenson's radical claim that our moral concepts are forward-looking is somehow misleading and should be reformulated in such a way that its equivalent will be resistant to Strawson's criticism. We have just considered how such reformulation could be done. But all these misleading claims do not undermine our general statement to the effect that acceptance of the simple-compatibilist paradigm<sup>25</sup> that makes up a hard-core of Stevenson's account is much more rational than acceptance of the incompatibilist paradigm.

## References

- Campbell, C. A. (1951), 'Is "free will" a pseudo-problem?', *Mind* 60.
- Frankfurt, H. (1986), 'Freedom of will and the concept of person', in J. M. Fisher, ed., *Moral Responsibility*, Cornell University Press, Ithaca and London, pp. 65–80.
- Hume, D. (2004), *An Enquiry Concerning Human Understanding*, Courier Dover Publications, Mineola, New York.
- Lakatos, I. (1978), 'Falsification and the methodology of scientific research', in J. Worrall and G. Currie, eds, *Methodology of Scientific Research Programmes. Philosophical Papers 1*, Cambridge University Press.
- Moore, G. E. (1912), *Ethics*, Williams and Norgate, London.

<sup>23</sup> See Strawson (1982), 70.

<sup>24</sup> See Strawson (1982), 78–9.

<sup>25</sup> By simple-compatibilist paradigm is meant a set of statements that entail free will is a pseudo-problem and responsibility is to be explained in terms of controlling one's behavior. This view was characteristic for Hume, Schlick, Stevenson etc.

- Schlick, M. (1978), 'Wann ist der Mensch verantwortlich?', in U. Pothast, ed., *Freies Handeln und Determinismus*, Suhrkamp, Frankfurt/M.
- Stevenson, C. L. (1938), 'Ethical judgments and avoidability', *Mind* 47, 45–57.
- Stevenson, C. L. (1944), *Ethics and Language*, Yale University Press, New Haven.
- Strawson, P. F. (1982), 'Freedom and resentment', in G. Watson, ed., *Free Will*, Oxford University Press, pp. 59–80.
- Van Inwagen, P. (1986), 'The incompatibility of responsibility and determinism', in J. M. Fisher, ed., *Moral Responsibility*, Cornell University Press, Ithaca and London, pp. 241–50.