Excuses and exemptions

Miranda del Corral
CONICET, Argentina
miranda.delcorral@gmail.com

Abstract

Excuses and exempting conditions aim to mitigate responsibility. This paper proposes a distinction between excuses and exemptions in terms of the distinctive kind of judgement each of them aims to respond. I argue that exemptions affect the explanatory relevance of the accused, while excuses fully or partially justify her, by affecting the evaluative claim involved in responsibility attributions. This distinction supports the claim that attributing responsibility is a two-step process, each of them corresponding to a different kind of responsibility—agential and moral—whose attribution is guided by two different although related cognitive and argumentative tasks: explaining an outcome, and evaluating its moral significance.

Keywords: moral responsibility, explanatory judgements, excuses, exempting conditions.

Resumen

Las excusas y las condiciones eximentes tienen como finalidad mitigar la responsabilidad. Este artículo propone una distinción entre excusas y eximentes basada en el tipo distintivo de juicio que cada una trata de responder. Argumento que los eximentes afectan la relevancia causal del acusado, mientras que las excusas lo justifican total o parcialmente, porque afectan al juicio evaluativo implícito en las atribuciones de responsabilidad. Esta distinción apoya una concepción de las atribuciones de responsabilidad como un proceso en dos etapas, donde cada

una de ellas corresponde a un tipo diferente de responsabilidad (agencial y moral), cuya atribución está guiada por dos tareas cognitivas y argumentativas diferentes, aunque relacionadas: explicar lo que produjo una situación, y evaluar su significación moral.

Palabras clave: responsabilidad moral, juicios explicativos, excusas, condiciones eximentes.

0. Introduction

The concept of responsibility plays an important role both in Ethics and in the Philosophy of Action,[1] often with very different meanings. Despite its pervasiveness, the attempts to bring together the agential and moral aspects of responsibility —or to find out whether it is possible, or fruitful, to do so —are scarce. During the last decade, though, the relation between moral and intentional judgements has received increasing attention, specially within the fields of Experimental Philosophy and Metaethics. The aim of this paper is to approach the problem of the relation between agency and responsibility through an analysis of the conditions that mitigate responsibility, that is, excuses and exempting conditions, such as duress, insanity, overlapping of moral duties and provocation, amongst others.

The idea that different mitigating conditions achieve the same end —to relieve the agent from her responsibility— through different means is not new. Austin proposed the following distinction between *justifications* and *excuses*: "In the one defence, briefly, we accept responsibility but deny that it was bad: in the other, we admit that it was bad but don't accept full, or even any, responsibility"(Austin, 1956, 2). By justifying our actions, we attack the claim that what we did was wrong. By excusing our behaviour, we explicit some conditions that mitigate our responsibility. Some excuses, such as insanity, exempt the agent from any responsibility. Exemptions would work as relievers of the agents' duties: "[e]xcusing someone in this sense amounts to exempting him from what would otherwise be a requirement, or at least an expectation" (Baron, 2007, 32).

---

[1] The concept of responsibility is also central in legal and epistemic contexts; however, for the sake of simplicity, I will leave legal and epistemic responsibility aside.

From a different tradition, Strawson (2008, chap. 1) claimed that there are two different kinds of pleas that aim to modify the feeling of resentment that is the basis for holding someone responsible. The first kind, which would correspond to *excusing conditions* (Watson, 1987), are meant to point out to certain conditions that make the agent's *actions* an inappropriate target of reactive attitudes. The second kind of pleas aims to show that the *agent* herself is not an appropriate target of such attitudes. Within this category, Strawson makes a distinction between 'transitory' (such as acting under post-hypnotic suggestion) and 'enduring' (such as being a child, or being deranged) *exempting conditions* (Campbell, 2005). Wallace (1994) adopts and develops Strawson's categorisation, and argues that exemptions, unlike excuses, affect the agent's powers of reflective self-control.

My aim in this paper is twofold. On the one hand, I want to suggest that mitigating conditions may be divided in excuses and exempting conditions by attending to the distinctive kind of judgement each of them aims to affect. Attributing moral responsibility consists both in making an *explanatory* claim about the outcome, which can be affected by exempting conditions, and in making an *evaluative* claim about the moral status of the agent or her actions, which can be affected by excuses. On the other hand, I will argue that this distinction can shed some light on the debate about responsibility attributions: the two-staged model of responsibility it supports explains some of our intuitions about the kind of judgements involved in attributing responsibility.

The structure of this paper is as follows. The first section analyses the causal link between the agent and the outcome that, intuitively, must hold in order to attribute responsibility to the agent for that outcome. *Agential responsibility*, I will suggest, is attributed when the outcome is explained in terms of the agent's authorship. Exempting conditions diminish or nullify agential responsibility because they affect the explanatory relevance of the accused. On the other hand, excuses, as it will be argued in the second section, aim to block the transition from agential to *moral responsibility*. In the third section, I argue that the distinction proposed presents an argumentative view of responsibility attributions, and it supports a two-staged account of responsibility. In the last section, I will discuss the distinction between normative and empirical models of responsibility, and I will present some of the experimental results within the recent field of experimental philosophy.

1. Exemptions and agential responsibility

It is a strong and widespread intuition that, in order to make agent A responsible for outcome O, there needs to be a causal relation between A and O (Sartorio, 2007; Moore, 2009). When we hold someone responsible for a plant's death, we also express a causal judgements: that the plant died *because of* that agent's actions or omissions. That is, we relate the agent to the outcome in a causal explanation.

Causal efficacy, however, is too broad to constitute the sole basis for holding someone responsible, for there are many things that we cause, and we are not responsible for all of them. Furthermore, omissions (Sartorio, 2004), causally overdetermined outcomes (Funkhouser, 2002), and negligible contributions to a collectively produced outcome (Mellema, 1985) challenge the view that causal efficacy is required for moral responsibility.

A solution to the problem of the relation between causation and responsibility is to consider that the appropriate relation between an agent and an outcome is not necessarily that of causal efficacy, but of *explanatory relevance* (Beebee, 2004; Björnsson and Persson, 2012). What is important is the role played by an agent in the causal explanation of the outcome for which responsibility is being attributed. A causal explanation consists, in general, in answering a question about why something is the case. A correct explanation does not require to mention all the causal relations and factors into play; some of these factors are selected and highlighted, depending on their relevance for the context of the explanation. Relevance is given by the context against which the explanation is required (Menzies, 2004). Background assumptions discriminate the important or relevant causes from other candidates, through contrasting its role in the production in the outcome, and through evaluating alternative outcomes. Background assumptions include both the inquirer's and the explainer's beliefs, presuppositions, and expectations. They function as a *ceteris paribus* clause, stressing the relevance of certain factors when all the other background conditions remain constant (Schweder, 1999). The practice of asking for and offering explanations is contextual to the background assumptions against which the explanation is required. Thus, an event has only causal relevance compared to other events; as Hitchcock claims, "relations of positive

or negative causal relevance only hold relative to specific alternatives" (Hitchcock, 1996, 402). The alternatives against which the causal relevance of the agent is assessed include those made up by normative expectations, such as a standard of reasonable person (Bok, 1998). This is why it is possible to be responsible for an outcome even in cases in which the outcome was not intended nor foreseen by the agent. When an agent acts negligently, she has not acted as a "reasonable person", this is, with reasonable care (Keating, 1996); this violation of the normative standard of a reasonable person makes her actions relevant in the explanation of the outcome. In other words, our beliefs about what an agent should have done (because it is reasonable to do so) are part of the basis for evaluating what role that agent has played in the production of the outcome.

Thus, when we hold an agent responsible, we take the agent to be part of the explanation of why the outcome is the case. This is why the concept of 'agential responsibility' gathers in a more suitable way the explanatory relation between an agent and an outcome, instead of that of 'causal responsibility'. Agential responsibility broadly corresponds to *authorship*: to attribute authorship of an action or state of affairs consists in explaining that action or state of affairs appealing to the author's reasons, those for which the agent acts, and that the agent acknowledges as *her* reasons (Frankfurt, 1987; Moran, 2001).

Agential responsibility, although based on normative considerations, is a non-moral kind of responsibility, that is, it is conceptually independent[2] from the moral evaluation of the outcome. It merely reflects that the agent is the author (or co-author) of the outcome. For example, I know that riding on a rollercoaster makes me feel queasy; and despite of that fact, I like doing so. Last weekend, I went to an amusement park and decided to go on the rollercoaster, and as a result I felt (unsurprisingly) dizzy for a while. I am agent-responsible for my dizziness, insofar it can be explained through my choices and actions.

---

[2]   However, causal judgements may affect moral judgements, and vice versa, as experimental philosophers suggest. Insofar as explanatory relevance depends on normative considerations about how the agent should have acted, there is room for moral considerations to play a role in that normative background against which the explanation is required or offered. I will discuss some experimental results in Section 4.

But no moral consideration is here at stake: my action was morally neutral.

Certain mitigating conditions attempt to reconsider the explanatory link between the agent and the outcome. I shall call those conditions *exempting*. They mitigate, or even undermine, agential responsibility, and they do so by making it inappropriate to explain the outcome in terms of the agent's authorship. They modify the background assumptions against which the explanation is required and offered, in a way that the relevance of the agent is diminished or nullified. Exempting conditions often provide an alternative explanation of the outcome. They are not an all-or-nothing matter: they can just put forward that a previously ignored (or understated) factor plays an explanatory role in the production of the outcome, hence diminishing the relevance of the agent.

There would be two main kinds of exempting conditions: those that interfere with the accused agential capabilities (either internal or external), and those that point out that certain external conditions are relevant in the explanation of the outcome, either nullifying the agent's role, or diminishing its importance.

The first kind of exempting conditions has been largely analysed in the philosophical literature on moral responsibility, because they refer to the conditions an agent must fulfil in order to be held responsible, such as freedom conditions, selfcontrol, reason-responsiveness, amongst others (see, for example, Pettit, 2007; Braham and Hees, 2012). The lack of one or more of these conditions can be presented as exempting from responsibility. While it may be correct to assert that this is so because these conditions are necessary for (moral, rational) agency to be exercised, my aim is to show that they exempt the agent because they play a relevant role in the explanation of the outcome, to the point of diminishing or nullifying the explanatory role of authorship. For example, a person under hypnosis will act as commanded no matter what reasons she might have for or against acting in such way; thus, she is not exercising her agency, or put otherwise, her agential capabilities are not explanatory. She is thus partially or totally exempt of responsibility. The difference between partial and total exemptions is a matter of the explanatory role of the agent, given the exempting conditions that hold: it may be the case that she does not play any explanatory role, or that her role is diminished. For instance, the degree up to which alcohol (or other drugs) make the agent exempt will depend on the amount of

alcohol consumed, and on the causal effects of the alcohol on the agent's capabilities.

Not all the exemptions require that the agent lacks one or more of the conditions to be held responsible. A second kind of exemptions are those cases in which the agent meets the criteria above, and nonetheless it is not appropriate to explain the outcome in terms of authorship alone. Greene and Cohen (2004) illustrate this possibility through the following example:

> Let us suppose, then, that a group of scientists has managed to create an individual —call him 'Mr Puppet'— who, by design, engages in some kind of criminal behaviour: say, a murder during a drug deal gone bad. The defence calls to the stand the project's lead scientist: 'Please tell us about your relationship to Mr Puppet...'

> "It is very simple, really. I designed him. I carefully selected every gene in his body and carefully scripted every significant event in his life so that he would become precisely what he is today. I selected his mother knowing that she would let him cry for hours and hours before picking him up. [...] At any rate, my plans for him succeeded, as they have for 95% of the people I've designed. I assure you that the accused deserves none of the credit."

> What to do with Mr Puppet? Insofar as we believe this testimony, we are inclined to think that Mr Puppet cannot be held fully responsible for his crimes, if he can be held responsible for them at all […] [G]iven the fact that forces beyond his control played a dominant role in causing him to commit these crimes, it is hard to think of him as anything more than a pawn (Greene and Cohen, 2004, 1780).

Mr Puppet does not lack any control or reason-responsiveness conditions. He is not either under the *direct* control of the group of scientists; instead, they have set up all the variables (with an error margin of 5%) for shaping his identity. The actual outcome is that Mr

Puppet has committed murder in a drug deal gone bad. However, knowing the prior conditions makes a difference in the explanation of the murder. Was the murder planned, it would be more easily explained by appealing to scientists' plan, rather than Mr Puppet's motives to commit the murder, because the explanation of his motives would point to the scientists' plan as well. On the other hand, if Mr Puppet acted unexpectedly, thus changing the planned course of action, intuitively, he could be held responsible for the murder.

Note that Mr Puppet's case does not necessarily entail a deterministic conception of agency. It could be well the case that Mr Puppet's actions could not be predicted with a 100% accuracy because of his freedom to choose. What this example points out is that the more we know about the causal chain leading to an agent's choices and actions, the more complete their explanation will be. Adding more additional factors to a causal explanation usually diminishes the relevance of each factor.

Hence, both the agent's capabilities and the explanatory relevance of other causal factors may be presented as exempting conditions. In both cases, those conditions aim to modify the background assumptions against which the explanation of the outcome for which responsibility is attributed. Those assumptions are conformed by the empirical and normative expectations concerning how the world and the agent will, and should, act.

There are two cases, though, that challenge this view. The first of them is drunkenness: a drunken person usually does not display a high level of self-control. Although she does not meet the conditions for being held responsible, it is usually considered that she is instead responsible for getting drunk in the first place: in the words of Sir Francis Bacon, "if a mad man commit a felonie he shall not lose his life for it, because his infirmity came by the Act of God; but if a drunken man commit a felonie, he shall not be excused because his imperfection came by his owne default"(Bacon, 1630, 34). The second problematic case is forgetting: it is non-exempting, but apparently inevitable. Suppose that Ann has promised to water Bob's plants; unfortunately, she forgets about her promise, and as a result, Bob's plants die. Ann's forgetting can be hardly seen as voluntary: having made that promise simply did not come to her mind.

Those two cases (drunkenness and forgetness) indicate that, in our cultural context, the usual background assumptions include a normative standard of conduct regarding the use of alcohol (or other

drugs) and a normative expectation not to forget one's duties. The normative expectations concerning the use of alcohol include the belief that people should foresee what they are capable to do while drunk, and avoid drinking if they believe they will lose control. On the other hand, concerning forgetfulness, we expect from people to remember their duties: if one makes a promise, not only one should fulfil it, but also take the necessary steps not to forget it. However, in both cases, it seems more appropriate to claim that, even if drunkenness and forgetfulness partially or totally make the agent exempt, she is nonetheless responsible for getting drunk in the first place, and for failing to remember her duties.

Therefore, both the agential conditions to be held responsible, as well as the standards for tracing the agent's actions and values, belong to the set of background assumptions against which the explanation of any outcome is required and offered. Furthermore, these assumptions include normative expectations about what should or should not be the case. Whether a circumstance is considered as exempting or not will depend precisely on the context of the explanation.

2. Excuses and moral responsibility

While exempting conditions aim to influence the explanatory judgement about the outcome, excuses are offered to affect the evaluative judgement about the agent who is held responsible. If the kind of responsibility at stake is moral, the evaluative judgement will be moral: it is asserted that what the agent did was morally wrong (or right). While agential responsibility entails an assessment of the agent and the outcome in the light of some normative standard, as that of a "reasonable person", moral responsibility entails an assessment of the agent in the light of moral standards. Moral expectations are the basis for attributing moral responsibility. Excuses aim to put forward certain conditions that may modify the moral expectations over the agent.

Thus, agential responsibility conceptually precedes moral responsibility. If an agent is not agent-responsible for an outcome, she cannot be held morally responsible for that outcome. In other words: if the agent is exempt, there is no basis for evaluating the moral status of the agent with regard to that particular outcome. In fact, excuses aim to block the transition from agential to moral responsibility, or, in Duff's terms, from responsibility to moral criticism or blame (Duff, 2007). That

is, the agent admits her authorship, but presents an argument against the claim that she acted wrongly:

> To excuse my action is to admit I had conclusive reason not to act as I did —that I acted either against a categorical, indefeasible reason, or against the balance of reasons; but to plead that I could not reasonably have been expected to act in accordance with either that categorical reason or the balance of reasons— which is to say, since the expectation that is involved here is clearly a normative one, that I cannot reasonably be condemned for failing to act thus. To offer an excuse is thus to admit responsibility, but deny liability: I admit to committing an action for which I must now answer, but seek to block the otherwise legitimate transition from responsibility to liability (liability, in this context, to moral criticism or blame) by offering an exculpatory answer (Duff, 2007, 53).

It might be possible, in principle, to excuse someone for a praiseworthy outcome. I will turn to the relation between exemptions, excuses and praise in the last section; for now, I will assume that excuses are put forward to affect negative moral judgements.

Excuses have justificatory force, either partial or total. An agent is fully justified when her actions cannot be considered wrong, and she is partially justified when her actions respond to some moral standards (although she has violated other standards). Full justification is exculpatory; partial justification is mitigating (Franklin, 2013).

The difference between full and partial justifications is that the former, but not the latter, denies moral responsibility: it is possible to explain the outcome in terms of the agent, but what she did does not constitute a violation of a moral standard. Full justification requires some sort of entitlement to act, so the action is not wrong for that agent (Botterell, 2009). For example, suppose that a doctor is administering a patient morphine to relief her from her suffering, and knows that her patient has a high probability to die as a consequence of the drug. When attributed responsibility for the death of the patient, she can argue that she had the permission (due to her medical authority and to the patient's will) to do it, and that it is not wrong to administer morphine to terminal patients. Thus, the doctor is justified in administering

morphine. She is agent-responsible for the patient's death; however, the doctor did not acted wrongly, for it was morally permissible for her to do it.[3] Justifications usually argue against the moral expectations that the attributer of responsibility has. In the example above, justifying to administer morphine to a terminal patient implies to claim that it is normatively expected from the doctor to do so —perhaps through arguing the normative reasons to guarantee a terminal patient a painless death.

Overlapping moral duties usually fully justify the agent, if those duties entitle the agent to act as she did. Suppose that Ann promised Bob to water his plants while he is on holiday. Bob goes on vacation and, two days later, Ann's mother falls severely sick, and her life is at risk. Whilst Ann's promise does not cease to be in force, she is morally entitled to unfulfil it to care for her mother.

An excuse, on the other hand, may offer a partial justification: the accused is, in this case, agent-responsible for that outcome, and that outcome or the actions leading to the outcome are morally wrong; however, there are (some) moral standards that the agent has respected. In this sense, partial excuses acknowledge that the agent has violated some moral expectations, but attempt to provide reasons that make the conduct reasonable. Following our example above, suppose that Ann has promised Bob to water his plants. When Bob comes back from vacation, his plants have dried —Ann did not fulfil her promise. She may try to excuse her behaviour by explaining why she did not water the plants. For instance, she might allege that she had a very busy week and that she left work late, so going to Bob's house became an effortful task. She might have tried to delegate her duty to someone else, without success. Hence, while it is true that she broke a promise, she had reasons to do it. Whether those reasons are good reasons will depend on the judging agent: while, to some, Ann would not be excused at all, to others she is partially justified, or even fully justified. Practical justification depends, as in the case of agential responsibility, on the empirical and normative standards against which the moral evaluation is made.

Responsibility attributions, then, entail both an explanatory and an evaluative judgement. Even if both judgements are often made against the same background assumptions, they should remain independent,

---

[3]   I am aware that this example is controversial. I assume here that the medical procedure described in the example is morally correct.

for they have different aims: while explanatory judgements aim to figure out why something happened, moral evaluative judgements aim to assess the moral significance of the agent's role in the production of the outcome. Although background assumptions are socially constructed (such as the standard of the 'reasonable person') and shared amongst the members of a community, judgements are individually made —this allows for the possibility of disagreement concerning attributions of responsibility. In fact, the practice of offering excuses and exempting conditions may be better understood as an argumentative practice: after all, excuses and exemptions play the role of *justificatory* and *explanatory* reasons.[4]

3. Responsibility as a two-step process

So far, I have argued that while both excuses and exemptions have the capacity to modify the degree of responsibility attributed, they aim at different stages of the process of attributing responsibility. On the one hand, exempting conditions diminish or eliminate the relevance of the role of the agent in the explanation of the outcome for which responsibility is being attributed. An agent's obsessive-compulsive disorder plays a more relevant role in the explanation of why this agent washes her hands every ten minutes than this agent's reasons and beliefs —such as her belief that her hands are full of bacteria. That is, she does not display full authorship, for her washing her hands is not responsive to her reasons. On the other hand, we offer an excuse to show that, even if the outcome is explainable through our authorship, our choices are fully or partially (morally) justified by our reasons. It may be correct to explain a dog's death through appealing to the treatment that a veterinary surgeon has applied; however, the vet may offer different excuses to affect her responsibility. She can adduce that the dog was in a terrible condition and that it was morally permissible (or even mandatory) to put a painless end to its suffering; in this case, she is offering a full justification. Not knowing that the dog was allergic to that particular treatment, on the other hand, may fully or partially justify her, depending on whether it was reasonable for her to know

---

[4]   This is a standard distinction in the philosophy of action; see, for an overview of the kinds of reasons, Álvarez (2009).

about that allergy. Moral responsibility always requires an evaluation of the agent's reasons, and excuses aim to intervene in this evaluation.

The two tasks involved in the process of attributing responsibility —explaining and evaluating— are often done simultaneously. Certain defences could be considered both an excuse and an exempting condition. For instance, duress affects both the explanatory claim and the evaluative claim. If I have been threatened to rob a bank, the threat is relevant in the explanation of the robbery of the bank, to the point it diminishes the explanatory relevance of my reasons for robbing a bank, and hence my authorship. Even if it did not make me fully exempt (because, let us say, the threat consists in the publication of a compromising photograph of mine, so not robbing the bank was still available as an option for me), the threat may also justify (again, either fully of partially) my actions, that is, robbing the bank.

Now, what are the implications of the proposed distinction for a theory of responsibility? While this distinction does not explicitly favour a particular theory of responsibility, it does support an account of what responsibility attributions consist of. First, as I have argued throughout this article, attributions of responsibility involve two distinct kinds of judgements: an explanatory one (the outcome happened because of the accused), and an evaluative one (what the accused did was morally wrong). So, what kind of judgements are excuses and allegations of exempting conditions themselves? Insofar as exempting conditions are adduced to mitigate the explanatory role of the accused, they are presented in explanatory judgements as well. Excuses, on the other hand, aim to justify the accused, and therefore they are evaluative judgements. Against the accusation of having acted wrongly, the accused may try to justify her actions, that it, to challenge the wrongness of her actions. This suggests that holding someone responsible has an argumentative aspect: ultimately, attributing responsibility is way of demanding answerability from the accused. In fact, it is reasonable to claim that blame has always to be open to refutal, for the accuser might be wrong. So, the distinction presented here supports a dialogical and social view of responsibility. Hence, the distinction is not meant to explain the difference between *being* excused and *being* exempt, or between being *appropriately* excused or exempt. It is a distinction between two ways of responding to responsibility attributions. The problem of the relation between being responsible and being held responsible (Smith, 2007) remains open. In order to know what being excused (or exempt) or

being appropriately excused (or exempt) amount to, we would have to tackle the issue of what being morally wrong (or explanatorily relevant) (*simpliciter*) amounts to.

Second, this distinction between excuses and exemptions, that is, between the evaluative and the explanatory judgements involved in responsibility attributions, supports a two-staged model of responsibility attributions, first proposed by Ross, who suggests the following metaphor of responsibility as a *trial*:

> The connection of responsibility with a trial shows that to be responsible for something can mean basically two different things corresponding to the two steps in the trial: accusation and judgement. In the first place being the person who can, when the situation demands, be rightfully accused (required to answer, give account); secondly, being someone who also satisfies the conditions of guilt and can therefore be rightfully sentenced (Ross, 1975, 17).

To rightfully *accuse* someone, then, consists in finding out whether there is an agential relation to the outcome, and whether the agent is exempt. I have labelled this kind of responsibility *agential*. On the other hand, to rightfully judge someone involves an evaluation of the conditions of guilt; if those conditions apply, the agent is judged to have acted wrongly. Given certain circumstances, the agent might be fully justified and therefore agent-responsible, but not accountable. Excuses may also partially justify the agent: she may be accountable, but her moral responsibility may be diminished.

Third, it is important to note that, in the analysis presented, excuses and exemptions do not aim to affect blame, that is, the practice of social and public punishment of the accused, either in the form of expressions of resentment, condemnation or any other act, directed to the accused, made in response to her actions.[5] Of course, it is possible to adduce certain conditions that aim to affect the response to some morally wrong action. In fact, apologies aim to diminish blame: they do not justify the accused, but acknowledge that she has acted wrong, and that she

---

[5]   Of course, if one maintains that blame consists precisely in the negative moral evaluation of the agent, then excuses do aim to affect it. I am now concerned with the social practice of blame or praise.

already experiences a negative feeling about the impact of her actions (if the apology is sincere). Apologies, though, are not the only way of mitigating blame. Suppose that Bob finds out now that Ann lost a book he had lent her a long time ago. Even if he finds Ann's action (of losing the book and not telling him) morally faulty, it may not make sense to blame her for it now. On the contrary, other circumstances make it appropriate to blame the agent: if she has acted wrongly many times, she may receive a greater punishment than if she were acting for the first time. Excuses and exemptions may affect blame, but they do not show whether the agent deserves blame or praise.

4. Experimental and normative accounts of responsibility

The argument offered here is analytical: it aims to establish conceptual connections between attributing responsibility, offering excuses, and alleging exempting conditions. Those phenomena, though, have received increasing attention from an empirical perspective, specially within the fields of experimental philosophy (see Knobe et al., 2012 for an overview) and moral psychology (see Ditto et al., 2009). In this last section, I will discuss some of the empirical findings concerning the relation between causation and moral responsibility.

Before considering the main hypotheses tested by recent works on moral responsibility and causal judgements, I will briefly introduce some relevant differences between empirical research and conceptual analysis. Since the emergence of experimental philosophy, there is an ongoing intense meta-philosophical debate concerning the role of intuitions —particularly, the intuitions of *analytic philosophers*— in the explanation of human nature. Experimental philosophers claim that many philosophical questions can and should be addressed through empirical methods. Other philosophers, on the contrary, defend intuitions as a legitimate justification of philosophical claims (Kauppinen, 2007). Although I sympathise with this latter view, providing a defense would exceed the scope of this paper. But since this section examines the framework presented above in the light of some experimental results on attributions of responsibility, I need first to address the issue of how, in my view, empirical results and conceptual analysis relate to each other.

Analytical arguments, such as the one carried throughout this article, aim at conceptual refinement, at the establishment of explanatory hypotheses that may or may not be suitable for empirical testing, and

at the explicit statement of philosophical intuitions, amongst other
things. Experimental methods serve to test hypotheses, and to isolate
new phenomena, amongst other purposes. Experimental philosophy
has the goal of testing philosophers' intuitions —those that are used
to support an argument, to provide counterexamples, to suggest
thought experiments (Knobe and Nichols, 2013). In this context,
testing philosophers' intuitions amounts to checking whether those
intuitions are pervasive amongst other social groups. However, while
I acknowledge the scientific relevance of enhancing our knowledge of
people's moral intuitions, I also find philosophically and scientifically
valuable to suggest new conceptual frameworks and to construct new
normative models for understanding human nature.

Many, if not most facets of human nature are *normative*: examples
of normative phenomena are language and argumentation, reasoning
and knowledge, moral actions, legal institutions, social agreements and
collective actions. To say that they are normative is to say that there are
correct and incorrect acts within those spheres. We may argue cogently
or fallaciously, or we may incorrectly believe that the Earth is flat. We
may speak German poorly or proficiently, and we may be great or
terrible football players. We break promises, and we are (sometimes)
akratic. In order to know whether we ourselves, or others, are doing
something *wrong*, we appeal to *normative standards*. Those standards
aren't fixed: there are historical, cultural and even subjective differences
between standards. I will leave aside the question of whether those
normative standards are normative facts, objectively knowable. For the
sake of the argument, it suffices to acknowledge that those standards
exist—otherwise, it would not be possible to judge that some certain act
is incorrect.

The conceptual framework I have argued for in this paper includes
normative standards. On the one hand, I have argued that causal
explanations are required and offered in a context, which determines
why we attribute relevance to certain things and not to others. Normative
standards of all kinds are part of those background assumptions. On the
other hand, evaluative judgements are also made within a context that
gives saliency to some events, facts and properties. Moral evaluations are
specially, although not exclusively, made against moral standards. The
framework here presented is itself a normative model of responsibility:
it could, in principle, be used as a normative standard for attributing
responsibility.

This is so because normative models are prescriptive: they tell us how we should act, and serve to evaluate whether an act is correct or incorrect according to the standard. Descriptive models, on the other hand, aim at accurate representation. For example, normative models of rational choice prescribe how to decide —one should choose the option that maximises her utility. Descriptive models of rational choice examine the actual process of choice —very often, they greatly diverge from what is prescribed by normative models (Baron, 2004). The methodological problem lies then in how empirical findings ought to affect normative models: should they be abandoned, or modified, in the light of mismatching evidence? I will turn again to this point later.

I will now examine some of the evidence gathered by experimental philosophers and moral psychologists concerning the relation between causal and moral judgements, and I will argue that the evidence is *compatible* with a two-staged normative model of responsibility.

Evidence from a variety of experimental settings support the claim that normative considerations play a role in people's assessment of causal factors. Those experiments involve two or more similar narratives, which only differ in a certain normative aspect of the story: events and characters vary in their typicality, their moral status, their social role, or the desirability of their motives. Alicke (1992) pioneered the empirical study of how moral considerations influence causal judgements; he introduced the participants two scenarios:

> John was driving over the speed limit (about 40 mph in a 30 mph zone) in order to get home in time to...

Socially desirable motive

> ...hide an anniversary present for his parents that he had left out in the open before they could see it.

Socially undesirable motive

> ...hide a vial of cocaine he had left out in the open before his parents could see it.

Other cause

Oil spill. As John came to an intersection, he applied his brakes, but was unable to stop as quickly as usual because of some oil that had spilled on the road. As a result, John hit a car that was coming from the other direction.

Tree branch. As John came to an intersection, he failed to see a stop sign that was covered by a large tree branch. As a result, John hit a car that was coming in the other direction.

Other car. As John came to an intersection, he applied his brakes, but was unable to avoid a car that ran through a stop sign without making any attempt to slow down. As a result, John hit the car that was coming from the other direction.

Consequence of accident

John hit the driver on the driver's side, causing him multiple lacerations, a broken collarbone, and a fractured arm. John was uninjured in the accident.

Complete the following sentence: The primary cause of this accident was ___ (Alicke, 1992, 369).

Participants, Alicke found out, were more prone to identify John as the primary cause of the accident in the case in which he was motivated by a socially undesirable motive than when his desire was socially desirable. Alicke concludes that "with causal necessity, sufficiency, and proximity held constant, the more culpable act was deemed by subjects to have exerted a larger causal influence" (1992, 370).

A similar experiment was conducted by Knobe and Fraser (2008). They presented the participants the following vignette:

The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own.

The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly emailed them reminders that only administrative assistants are allowed to take the pens.

On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message... but she has a problem. There are no pens left on her desk (Knobe and Fraser, 2008, 443).

The subjects participating in the experiment were asked to state their degree of agreement with each of these propositions: "Professor Smith caused the problem" and "The administrative assistant caused the problem". Participants were prone to agree with the claim that Professor Smith caused the problem, while disagreed with the statement that the assistant caused the problem. However, Knobe and Fraser argue, both behaviours are equally frequent; the only difference stems from the different moral value of each action. Therefore, they conclude, moral judgements affect the process by which we make causal claims.

Knobe and Fraser's conclusion has been revised by Roxborough and Cumby (2009), who argue that the experiment lacks from a differentiation of a crucial factor to attribute a role in a causal explanation: the typicality (or atypicality) of events. The authors conducted another experiment, with the same scenario, but introducing a variation on how typical the rule-following and the rule-violation behaviour was. They conclude that the typicality or atypicality of a behaviour affect causal judgements. Furthermore, a variation in the typicality of the competing causes (in this example, the administrative staff behaviour) affects attributions of causal relevance to Professor Smith, even when the typicality of his behaviour remains constant.

Driver (2008) argues that the evidence obtained so far is compatible with a broader interpretation of the relation between causal and moral judgements:

When we make attributions of primary causation, that is, when we pick out the cause of an event among a nest of causal factors, it is quite true that we often rely

> on pragmatic and contextual considerations […] The
> identification in judgment of something's being the
> cause depends upon pragmatic factors and may well
> include moral judgments […] What we call a "cause"
> or "causal factor" in many cases depends on what we
> relegate to the background conditions of the event in
> question (Driver, 2008, 425).

Driver's claim, thus, is that normative considerations play a role in attributing causal relevance. The conceptual framework also acknowledges that this is so. As it has been suggested in the previous section, we may express an evaluative judgement and a causal judgement simultaneously —in fact, certain mitigating conditions, such as duress, are both excuses and exempting conditions. One and the same reason (that you acted under duress) is a reason for judging that other things had more explanatory saliency and a reason for changing our mind about the moral status of your actions and of yourself. Hence, a two-staged model of responsibility is *compatible* with the results obtained in the experiments discussed above. Now, is this all we can conclude?

The answer to this question depends on how we tackle the methodological and epistemic problem of the empirical (dis)confirmation of normative models, as well as on whether the framework presented above has an empirical interpretation, one that could be (dis)confirmed by the evidence. I will first address this second question.

The experiments above show that normative evaluations and causal explanations are two interrelated kinds of judgements, both of them relevant for attributing responsibility. In this paper, I have provided a different argument for the same conclusion. I have argued that mitigating conditions can be divided in two categories: excuses and exempting conditions. This distinction is based in the following intuition. Attributions of moral responsibility express an accusation: "What you did was wrong". To respond to the accusation, you may adduce that what you did is not wrong, or that it is less wrong than it initially appears to the accuser —you then offer an excuse. On the other hand, you may point out that you did not do what you are being accused of, or that there is more in the story of the causal production of the outcome that makes your intervention less relevant than it initially

appears to the accuser —you then offer an exempting condition.[6] So, the question is: if we take this distinction to be an empirical claim, do the experiments presented above (dis)confirm it? I doubt so. At best, as pointed out earlier, the evidence is compatible with this distinction.

An empirical research, though, may be performed in order to test people's intuitions about how different mitigating conditions work, and how they affect the evaluative and the causal judgements separately. It would be necessary to adapt the framework presented, and reformulate it so the main and auxiliary hypothesis are clearly stated. Without doubt, it would be scientifically and philosophically valuable to have a better knowledge of the actual practice of offering excuses and other mitigating conditions.[7] After all, attributing responsibility is often a social phenomenon: to attribute responsibility is to make people answerable, to make them respond to their own actions. Hence, in order to go one step beyond compatibility, the framework should be reformulated so the hypothesis are clearly stated; they could then be confirmed, or disconfirmed by the evidence.

Finally, the problem of the (dis)confirmation of a normative account of responsibility attributions through empirical evidence remains open. Solving it will require a deeper understanding of the epistemic properties of normative models. As a first approximation, we might say that if a model is impracticable, that is, it cannot be implemented in the actual object that is represented in the model, then there is good reason to reformulate or to abandon the model. Imagine a normative building evacuation model, which depicts how people ought to abandon a building in the case of an emergency. This models are used in the design of the building. Now, a building evacuation model whose instructions cannot be followed (for example: it assumes that people walk faster than they can) is a failed model and should be abandoned. In other cases, people tend to act contrary to what is prescribed in the model —think of the frequency of *ad hominem* claims. People are able

---

[6]   A third kind of response that I have mentioned, but that I have not developed here, is the one that aims to mitigate the acts of blame (or praise): you acknowledge full responsibility, and apologise, for instance.

[7]   The practice of justifying one's actions has in fact been long studied by social psychologists; see, for example, Weiner (1995). Of course, an empirical study of the excuses/exemptions distinction should take into account this previous experimental work.

to follow the model; they simply do not do it. Should the standard argumentative model be abandoned in the light of the pervasiveness of fallacious claims? The answer is negative: we have good reasons to keep those models, reasons that are independent from the frequency of deviating behaviour. For example, there are good reasons to maintain a truth-preserving argumentation model: if the premises are true, so is the conclusion. Those independent reasons have to do with the fact that one of the main goals of argumentation is to gain knowledge: we aim to know what to do, to know how things stand. The fact that truth is the aim of knowledge is a reason to create truth-preserving models of reasoning and argumentation, and this is a reason that is independent from the typicality of the deviation.

Similarly, there are good reasons to defend a normative model of responsibility attributions, reasons that are independent from the frequency with which people devate from the model —an example of deviation from this model, for instance, would be to make someone responsible while acknowledging that she had no causal influence on the outcome. Those independent reasons have to do with one of the main aims of attributing responsibility: the pursuit of fairness. It is unfair to attribute agential or moral responsibility for an outcome to someone who played no role in the causal production of the outcome, in the same sense that it would be unfair to attribute moral responsibility for a morally neutral outcome to someone who has not acted immorally. This is, I think, the main motivation to accept the normative framework presented in this paper.

Concluding remarks: What about praise?

I hope to have shown that a close examination of the different kinds of allegations an agent may present to mitigate her responsibility can shed some light on the explanation of the practice of attributing responsibility. I have narrowed my analysis to responsibility for wrongdoing, leaving praiseworthiness out of the discussion. The judgements involved in attributions of praise are the same as those involved in attributions of blame, that is, an explanatory and an evaluative judgement.[8] Mitigating

---

[8]   However, background assumptions for explaining a morally wrong outcome seem to be different from those that are present when explaining a morally good one; particularly, the relevance of the agent's intentions vary, as Knobe (2003) has show in a series of long-discussed experiments.

conditions for praise have received much lesser attention than their counterparts for blame. In principle, though, one may also try to affect the explanatory or the evaluative judgements involved in an attribution of praise.

Consider the following situation: Ann has consumed a large quantity of alcohol, and when walking back home, she sees a house set on fire; she then runs towards the house, enters it, and a few minuted later, she exits the house with a toddler in her hands, whom she has saved from the fire. Ann's actions are usually considered praiseworthy; however, we know that Ann would not have entered the house being sober. She felt overconfident because of the alcohol she had drunk. In this case, it seems reasonable not to attribute responsibility to Ann for rescuing the toddler, or at least to attribute only partial responsibility, for Ann having consumed alcohol plays an important role in the explanation of the outcome, and thus it would play the role of an exempting condition. On the other hand, there are cases in which one tries to affect the evaluative judgement involved in attributions of praise. Suppose that Ann has saved the toddler from the previous example without having consumed any alcohol: she just decided to enter the house and look for people in distress. When being praised for her action, she reveals that she works as a firefighter. Although having saved a life is no doubt a morally good action, Ann may argue that it was in fact her duty to enter the house and look for survivors. Here, the evaluative judgement is affected, and Ann's responsibility may be diminished. Of course, there is a lot more to say about mitigating conditions for morally good outcomes and actions; these two examples merely show that, in principle, it is possible to apply the distinction developed above to such kind of cases.

References

Alicke, M. D. (1992). Culpable Causation. *Journal of Personality and Social Psychology*, 63 (3), 368–378.

Alvarez, M. (2009). How Many Kinds of Reasons? *Philosophical Explorations,* 12 (2), 181–193.

Austin, J. L. (1956). A Plea for Excuses: The Presidential Address. In *Proceedings of the Aristotelian Society*, 57, 1–30.

Bacon, F. (1630). *The Elements of the Common Lawes of England*. 2003rd ed. New Jersey: The Lawbook Exchange, Ltd.

Baron, J. (2004). Normative Models of Judgment and Decision Making. In *Blackwell Handbook of Judgment and Decision Making*. D. J. Koehler, & N. Harvey (Eds.) (19–36). Malden MA: Blackwell.

Baron, M. (2007). Excuses, Excuses. *Criminal Law and Philosophy*, 1 (1), 21–39.

Beebee, H. (2004). Causing and Nothingness. In *Causation and Counterfactuals*. J. D. Collins, E. J. Hall & L. A. Paul (Eds.) (291–308). Mass: The MIT Press.

Berman, M. N. (2003). Justification and Excuse, Law and Morality. *Duke Law Journal*, 53, 1–77.

Björnsson, G., & Persson, K. (2012). The Explanatory Component of Moral Responsibility. *Noûs*, 46 (2), 326–354.

Bok, H. (1998). *Freedom and Responsibility*. Princeton: Princeton Univ Press.

Botterell, A. (2009). A Primer on the Distinction between Justification and Excuse. *Philosophy Compass*, 4 (1), 172–196.

Braham, M. & van Hees. M. (2012). An Anatomy of Moral Responsibility. *Mind*, 121 (483), 601–634.

Campbell, J. K. (2005). Compatibilist Alternatives. *Canadian Journal of Philosophy*, 35 (3), 387–406.

Ditto, P. H., Pizarro, D. A. & Tannenbaum, D. (2009). Motivated Moral Reasoning. In *Psychology of Learning and Motivation*. B. H. Ross (Ed.) 50, (307–338). Burlington: Academic Press.

Driver, J. (2008). Attributions of Causation and Moral Responsibility. In *Moral Psychology*. W. Sinnott-Armstrong (Ed.) 2, (423–440). Cambridge, Mass.: MIT Press.

Duff, R. A. (2007). Excuses, Moral and Legal: A Comment on Marcia Baron's 'excuses, Excuses.' *Criminal Law and Philosophy*, 1 (1), 49–55.

Frankfurt, H. (1987). Identification and Wholeheartedness. In *Responsibility, Character and the Emotions*. F. Shoeman (Ed.) (27–45). New York: Cambridge University Press.

Franklin, C. E. (2013). A Theory of the Normative Force of Pleas. *Philosophical Studies*, 1–24.

Funkhouser, E. (2002). Three Varieties of Causal Overdetermination. *Pacific Philosophical Quarterly*, 83 (4), 335–351.

Greene, J. & Cohen, J. (2004). For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359 (1451), 1775–1785.

Hitchcock, C. R. (1996). The Role of Contrast in Causal and Explanatory Claims. *Synthese,* 107 (3), 395–419.

Kauppinen, A. (2007). The Rise and Fall of Experimental Philosophy. *Philosophical Explorations,* 10 (2), 95–118.

Keating, G. C. (1996). Reasonableness and Rationality in Negligence Theory. *Stanford Law Review,* 48, 311–384.

Knobe, J. (2003). Intentional Action and Side Effects in Ordinary Language. *Analysis,* 63 (3), 190–194.

Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H. & Sommers, T. (2012). Experimental Philosophy. *Annual Review of Psychology*, 63, 81–99.

Knobe, J. & B. Fraser. (2008). Causal judgment and moral judgment: Two experiments. *Moral psychology*, 2, 441–448.

Knobe, J. & S. Nichols. (2013). An Experimental Philosophy Manifesto. In *Experimental Philosophy*. (3–14). Oxford, New York: Oxford University Press.

Mellema, G. (1985). Shared Responsibility and Ethical Dilutionism. *Australasian Journal of Philosophy,* 63 (2), 177–187.

Menzies, P. (2004). Difference-Making in Context. In *Causation and Counterfactuals*. J.D. Collins, N. Hall, & L. A Paul (Eds.) (139–180). Cambridge, Mass.: The MIT Press.

Moore, M. S. (2009). *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford: Oxford University Press.

Moran, Richard. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton NJ: Princeton University Press.

Pettit, P. (2007). Responsibility Incorporated. *Ethics*, 117, 171–201.

Ross, A. (1975). *On Guilt, Responsibility, and Punishment*. Great Britain: Univ of California Press.

Roxborough, C. & Cumby, J. (2009). Folk psychological concepts: Causation. *Philosophical Psychology,* 22 (2), 205–213.

Sartorio, C. (2004). How To Be Responsible For Something Without Causing It. *Philosophical Perspectives*, 18 (1), 315–336.

—— (2007). Causation and Responsibility. *Philosophy Compass,* 2 (5), 749–765.

Schweder, R. (1999). Causal Explanation and Explanatory Selection. *Synthese*, 120 (1), 115–124.

Smith, A. M. (2007). On Being Responsible and Holding Responsible. *The Journal of Ethics*, 11 (4), 465–484.

Strawson, P. F. (2008). *Freedom and Resentment and Other Essays*. New York: Taylor & Francis.

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

Watson, G. (1987). Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. In *Responsibility, Character and the Emotions*. F. Shoeman (Ed.) (256–286). Cambridge: Cambridge University Press.

Weiner, B. (1995). *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. New York: Guilford Press.

Westen, P. (2006). An Attitudinal Theory of Excuse. *Law and Philosophy*, 25 (3), 289–375.