A dialectical model of moral address

David Botting
IFILNOVA
Universidade Nova de Lisboa
26-C Av. de Berna 1069-061, LISBON, PORTUGAL
mephilsci@yahoo.co.uk

RESUMEN

Un punto de vista relativamente nuevo en la teoría de las actitudes reactivas interpreta comunicativamente las actitudes reactivas y sus expresiones, y alega que lo los agentes que son su objeto se caracterizan por su capacidad de participar en un intercambio dialógico de razones morales. La competencia moral se puede evaluar en términos de competencia lingüística, de forma que nuestras atribuciones de responsabilidad dependen en última instancia de lo que la gente puede decir aceptablemente en un cierto tipo de discurso moral.

McKenna (2012) ve todo el proceso, desde la comisión de la acción hasta la actitud después de reflexionar, por analogía con el despliegue de una conversación, y lo llama "intercambio de responsabilidad moral". Quiero discutir si la discusión crítica de los pragmadialécticos puede servir como modelo de diálogo para esa conversación. Si es así, las reglas de la discusión crítica determinan junto con los asunciones materiales iniciales si un agente puede ser considerado responsable por la otra parte de la discusión.

PALABRAS CLAVE: actitudes reactivas, conversación, modelo de diálogo, Michael McKenna, responsabilidad moral.

ABSTRACT

A relatively new view in reactive attitudes theory regards the reactive attitudes and their expression communicatively, and argues that what characterizes the agents who are their targets is a capacity for engaging in a dialogical exchange of moral reasons. Moral competence is evaluable in terms of linguistic competence, with our attributions of responsibility depending ultimately on what people can acceptably say in a certain kind of moral discourse. McKenna (2012) regards the entire process from the commission of the action to the attitude after reflection on analogy with the unfolding of a conversation, and he calls it a "moral responsibility exchange". I wish to consider whether the critical discussion of the pragma-dialecticians can serve as a dialogue model of this conversation. If it can, then the rules of critical discussion and the material starting-points together determine whether an agent should be held responsible by the other party in the discussion.
.

KEYWORDS: moral responsibility, reactive attitudes, conversation, dialogue model, Michael McKenna

1. THE BASIC PROGRAM: THE FORCE OF BLAME

Blame has a force and a warrant. Seeing blame as a speech act, we can identify the warrant with a condition of satisfaction of the blame and the force as its expected perlocutionary effects. The two things are interdependent, as we will see, for if the perlocutionary effects are potentially harmful to the addressee then we need a stronger warrant than if they were not. Furthermore, if we think that the harm is morally deserved then we need a different warrant and a different kind of justification for the blame altogether than if we do not think it is morally deserved. If we do not think that the warrant is sufficient then we cannot really blame; we may be able to perform the same act, but this is no longer analyzable as, and we would not see it as, blame, or at least not as *moral* blame.

The first two parts of this paper are concerned with unfolding what I call "the ordinary conception of blame" – the conception of blame we are committed to in our practices of blame and punishment and our second-order practices of justifying those first-order practices – and showing that: i) we do take the harm we potentially cause as morally deserved; ii) we do take the warrant required to justify this to involve moral reasons; iii) the best way of explaining how blame gets this force is given by reactive attitudes theory and, in particular, its conversational variant. I cannot in this place rule out the possibility that this ordinary conception is mistaken and that we should revise this conception, and with it (potentially but not necessarily) our moral practices; Strawson (2008) argues that we have a natural commitment to reactive attitudes such that our responsibility attributions do not come up for revision (though to say this is not to say that those reactive attitudes are always appropriate, but only that we would always think they were, that is to say, that the ordinary conception of blame is one that we cannot really let go of in practice no matter what our theoretical beliefs are), but defending this claim is well beyond the scope of this paper.

Taking his cue from Watson (2004), McKenna (2012: 3) makes the following program statement:

> Watson suggested that our practices of holding morally responsible and the expression of our moral emotions via those practices should be understood as forms of communication, expressions of moral demands and expectations, as well as expressions of our altered regard for those who fail to meet them. I agree, and I build upon Watson's proposal to develop not just an expressive and communicative theory of holding morally responsible, but a conversational theory in which holding morally responsible is understood as a stage in something

> analogous to an unfolding conversation. . . between competent speakers of a language, a dialogue between the morally responsible agent who is responsible, and those in the moral community holding her responsible.

By "holding morally responsible" McKenna means acts of overt and directed praise or blame. I am not concerned with praise, and "holding morally responsible" should henceforth be understood as implicitly completed by the words "for a wrong-doing" and associated with blame. Blame is overt when it is publicly manifested and directed when the one blamed is present and is the intended recipient of the manifestation.

By "moral emotions" McKenna means what Strawson (2008) calls the reactive attitudes. There are two kinds: participant (or non-detached) and moral (or detached). I will focus on the former, which are defined as the "non-detached attitudes and reactions of people directly involved in transactions with each other . . . such things as gratitude, resentment, forgiveness, love, and hurt feelings," (Strawson, 2008: 5) and as "natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions" (Strawson, 2008: 10-11). We adopt negative participant reactive attitudes (e.g., resentment) when we ourselves are the injured party and we think the injury is a manifestation of ill-will or indifference towards us, and by doing so we engage in a particular kind of interpersonal relationship with the target of our attitude (the wrong-doer).

By "should be understood as forms of communication" McKenna is giving overt, directed blame a particular force. By blaming we make those we blame the target of reactive attitudes and may (though not necessarily) alter our regard towards them in a way that reflects the impaired relationship we have with them and which for that reason causes harm to them, at least in so far as they can no longer feel entitled to expect, either now or in the future, the friendly reciprocal relations, or its benefits, that existed prior to the impairment. It seems to follow that for blame to be warranted the harm must be deserved, and this is so even if the blame is not intended as such to cause harm — we do not say that there was any such intention, still less that we blame *as a means to* causing harm, or *because* it will cause harm. It is a condition of satisfaction of the speech act of blame that we take it to be warranted in this way, where this is cashed out as the reactive attitude's being appropriate, though it is not a condition of satisfaction of blame, I will argue, that the speaker actually experience the reactive attitude. It is possible for a speaker not to experience the reactive attitude and still carry out the speech act of blame (though I will argue that there is still a reactive attitude expressed by the blame, but this is a reactive attitude that is attributed to the moral

community on whose behalf the speaker speaks).

There are linguistic indications that this accords with an ordinary 'folk' conception of blame. We do not typically blame by using "blame" as an illocutionary verb but by, for instance, expressing anger, resentment, or indignation towards the one responsible, and this is responsible for its force. The man who responds to his foot being deliberately trodden on with "I blame you," or announces his retaliatory strike with "I hereby blame you!" would be a very curious individual. The appropriate response would be, e.g., an expression of anger. We would expect "blame" to occur as an illocutionary verb more often than it does if there were not this intimate relationship between blame and the reactive attitudes.[1]

When we do use "blame" explicitly as a verb it is usually in order to apportion responsibility and is used evaluatively rather than to actually blame. Although the evaluative use is one plausible conception and use of blame, Watson and McKenna deny that overt, directed blame is merely the expression of an evaluation of moral competence, just as the grading of an English comprehension essay is the expression of an evaluation of linguistic competence; its force is not that of an assertive, of recording a fact in a moral ledger, and its warrant is not simply whether what has been asserted is justifiably believed to be true by the blamer.

Blame has a performative aspect that goes well beyond that of fact-stating discourse and seems to require a stronger warrant than that required for fact-stating discourse because it usually causes harm to the one blamed (when direct), or might do in the future (when not direct). We can conclude that ordinarily when we blame we do not state an evaluation, where this is indicated by the fact that we do not typically utter a sentence with the verb "blame." The conception of blaming as moral evaluation does not give our blaming the force that we ordinarily take our blamings to have or require the warrant that we ordinarily take our blamings to require, so this "ledger" view cannot be our ordinary conception of blame.

Although I deny that we use the illocutionary verb "blame" when we blame I do not deny that blaming, in its ordinary conception, is or can be viewed as an illocutionary and/or perlocutionary act. I only say that this speech act is not simply to convey information — it is at least among its perlocutionary effects to cause harm. Could we say, then, that blaming is just a kind of sanction, or an evaluation plus a sanction? As we will see in a while, what incompatibilists say (and Strawson agrees with them in this

---

[1] We might sometimes or even often drop the verb "blame", but only because we expect those we address to take it as understood.

regard) is lacking in the compatibilist Schlick/Smart view of blame as justified by its consequences is that blame lacks any moral dimension. The ledger view of blame, since it involves a moral evaluation, does give blame a moral dimension. Could we not achieve some kind of division of labor here?

No, because the kind of warrant required for a sanction is of a different kind than is implicit in our ordinary conception of blame, and the reasons for which we would think a sanction is warranted are of different kinds. The warrant for a sanction depends only on its effectiveness for social regulation and leaves out completely any issue about whether the harm is deserved. It is a purely instrumental, practical, forward-looking attitude towards the consequences of blame that sees moral practices purely as a means of social regulation. The kind of compatibilist that Strawson (2008) calls the consequentialist, as represented by Schlick and Smart, effectively identify blame and sanction and must perforce say that this is all there is to blame; to the complaint that this is not ordinarily how we think of blame when we actually blame or the kind of justification that we take it to have, the consequentialist has nothing to say except possibly that our conception of blame is simply wrong and that reactive attitudes are an understandable but mistaken facet of human psychology. Against them, we might argue that this is a sense that the word "blame" might be given, but it is not "blame" as we ordinarily think of it when we engage in it because it leaves out something vital to blame, leaving as the only attitudes that we are warranted in taking towards the blameworthy the 'objective' attitudes which treat the blameworthy only as things to be managed, while reactive attitudes (in which the retributive attitudes are included) would be inappropriate and unwarranted.

What this way of justifying our moral practices seems to leave out is the notion that the harm is deserved; consequently, the conception of blame that might be associated with objective attitudes seems to have quite a different warrant again from both evaluations and from ordinary blame.[2] On this point Strawson agrees with the libertarians, saying at (2008: 4) that this kind of instrumentalist justification does not justify our moral practices "as we understand them" and that there is "something more" that the consequentialist (by necessity) has left out of his account. Strawson (2008) argues that this vital something that the compatibilist leaves out is the reactive attitude itself. Reactive attitudes are vital and essential to blame,[3] and it is reference to the

---

[2] The incompatibilist may even allow that our moral practices *could* be justified in this way and continue on as before, but when we *do* justify them in this way we *do not* justify their having any moral force and effectively deny that they have any moral force. This is not blaming as we ordinarily conceive it.

[3] There are a couple of alternative readings of this thesis. The straightforward one is simply that one

reactive attitudes that adds a backward-looking feature in contrast to the consequentialist's entirely forward-looking justification, and gives blame its distinctive force.

In summary, the conception of blaming as sanctioning does not assign to our blaming the force or warrant that we ordinarily take our blamings to have; this conception may be adequate for blame as a perlocutionary act (for we may continue in what we actually do and the moral practices we actually have and may be warranted to do so even if we do not think that desert is necessary for moral responsibility), but not for blame as an illocutionary act. When we do justify them in this consequentialist way we do not conceive of our practices as having any moral force, for the justification is not of a kind that can justify its having any such force, since it is not a moral justification; hence, since we may still insist that as we ordinarily conceive it blame has a moral force and this force must be justified by a moral warrant, whoever asks us to believe that the non-moral blame is the only applicable kind – and consequently that the instrumentalist/consequentialist kind of justification he is prepared to give us is persuasive and the only kind of justification that can reasonably be expected – is asking us to conceive of and justify the practice differently from how we ordinarily conceive and justify it. Blame justified by its consequences is not *moral* blame and is not *morally* justified, whatever other virtues it may have. When we blame in this way we do not (and cannot without conceptual inconsistency) count our blame as "moral" or describe it as "moral" to ourselves or another, for our blame would not carry the implication that the one blamed deserves to be blamed; it is this implication that the moral warrant must justify, and it is this that distinguishes the consequentialist kind of justification from a distinctively moral kind.

It is the illocutionary act of blame that we are really interested in and that is a form of communication with the blameworthy. The consequentialist says not to worry because the objective attitudes that are appropriate and are warranted can provide the basis of our moral practices as they already are, forcing us to change not what we do

---

cannot blame without having the reactive attitude. Yet I doubt whether this is correct. Rather, it should be read as saying that reference to the reactive attitudes is ineliminable from an analysis of blame, just as reference to subjective colour experiences is ineliminable from the analysis of colours without it being the case that a particular subject must have a particular experience. This seems consistent with Strawson's (2008: 23) comment that "the making of the demand is the proneness to such attitudes." To blame is to make a moral demand and gets its force from its conceptual links to reactive attitudes even in the event of their absence. Similarly, Wallace (1994) argues that a judgment that reactive attitudes are apt will suffice and that one need not actually have those attitudes. Thus, it is an objective property of the blameworthy act that the correct response to it is resentment (indeed, this a conceptual truth about being blameworthy), and we can be in a position to judge that this is the correct response even if we do not actually feel resentment.

but only about how we feel about them and justify them. I respond that even if we do not act differently this is to change blame after all or to regard it only as a perlocutionary act.

We are in a position now to say something about the features the ordinary conception of blame should have. It must have a backward-looking element — it must be a warranted response to a violation of a moral demand. More particularly, it must be warranted as an illocutionary act. That is to say, it is still warranted even when it does not give rise to its usual perlocutionary effects, since we do not seem to be adopting a different conception of blame when we use the same speech act in cases where it is not directed, that is to say, when the intended target of our blaming is not available to be blamed and no harm ensues from the blame and hence, by any practical standard, our blaming seems to be a pointless gesture. There is a tricky intermediary case to be considered too that Sher (2012) calls "the problem of the stranger" when the morally blameworthy agent has no relationship to the one she injures, as for instance in a chance, never-to-be-repeated encounter. This seems to be a constraint on whatever we might suggest provides this vital something, and seems to caution against views that rest too heavily on interpersonal relationships between the blamer and the one blamed and the expectations they have with regards to each other's behaviour. It is because it avoids problems like this that a dialectified version of the reactive attitudes theory seems particularly useful, as will hopefully become clear.

Watson's innovation, extended by McKenna in the excerpt above, is to give the force of blame a communicative function Watson calls *moral address*, seeing the interpersonal transactions involved in participant reactive attitudes on the model of a dialogue. The backward-looking feature is publicly manifested in the speech act complex by linguistic and conversational mechanisms without necessarily being made explicit. I will show later that this is because it contains a reference to a moral demand/expectation that has been violated.

While satisfactory in terms of providing a backward-looking aspect, a communicative theory seems to be open to an immediate objection that we have just mentioned, which is precisely that it seems to work only for directed blame; *directed* blame seems to have this kind of communicative function, and it is essential to the warrant for directed blame that it does, but when the one to be blamed is not present and thus cannot be communicated with, blame seems to be unwarranted. McKenna's response to this kind of concern is to take directed blame as the paradigmatic case and claim that what is true of this case is true of the other cases of blame to the extent that

they resemble the paradigmatic case. His aim is not to define necessary and sufficient conditions that will cover all possible cases of blame (McKenna, 2012: 174-78).

I wish to take a different course: by replacing the attitudes with the dialectical notion of a commitment and noting that this allows there to be a commitment with analogous force and content as the reactive attitude even if there is no reactive attitude, the account I will offer will move McKenna's theory closer to Wallace's (1994) view that one does not need to have a reactive attitude in order to hold someone responsible but only a judgment that a reactive attitude is apt. I will be coming back to this and claiming that this avoids some of the problems that an account purely in terms of attitudes, or one like Wallace's in terms of attitudes and judgments, would face. This must wait, however. In the next section I will be discussing the warrant of blame.

2. THE BASIC PROGRAM: THE WARRANT OF BLAME

Although the reactive attitudes are emotions, and it is important that they be emotions in order to provide blame with its force, it is equally important that these emotions are under rational control such that they can be and actually are withdrawn under certain conditions where they seem inappropriate. Strawson's aim in his paper is to show that belief in determinism is not such a condition and could never, by itself, render our reactive attitudes inappropriate (Strawson, 2008: 7-10). Determinism is not our issue here, however.

There are two types of conditions: *excuses* and *exemptions*. Together these should tell us when reactive attitudes are warranted, that is to say, tell us when it is appropriate to hold an agent responsible.

Excuses function by showing that an action did not manifest ill-will. One kind of reason that the norms of appropriateness are sensitive to, then, are those concerned with quality of will. Importantly, an agent excused in this way is still held to expectations and as one for whom reactive attitudes are apt (Strawson, 2008: 7-8).

Exemptions function by showing that the agent's agency is too diminished by their incapacity to understand expectations or norms or to regulate themselves by those expectations or norms for it to be appropriate to hold them to those expectations or norms; they are not apt subjects for reactive attitudes (because not candidates for the kind of reciprocal relationships that those attitudes make possible). A second kind of reason that the norms of appropriateness are sensitive to, then, are those concerned

with incapacity (Strawson, 2008: 8-10). Here the reactive attitudes should be replaced with objective attitudes, and our moral practices may or may not change with them.

Watson (2004) characterizes this difference in the following way: agents that are apt objects of the reactive attitudes are likewise apt for *moral address*, where this means that they can be challenged over their actions and can be expected to give a response, and it is lack of a capacity for *moral address* that unifies the exemptions as a class and explains why it would be inappropriate to hold exempt agents responsible. Those incapacitated in this way are unsuited to certain types of interpersonal relationships; in such cases expressing the emotion could not carry out its characteristic communicative function and the speaker, when he learns this, would be in the performative contradiction of attempting a communication that he knows cannot succeed. This is meant to be (although I do not think Watson succeeds in this for all the cases he considers) a conceptual truth about the illocution involved; it is not simply the fact that expressing the attitude will not bring about the desired effect that makes the attitude inappropriate, for that would amount to the consequentialist objection that in cases of exemption blame would be pointless, but that is not a moral objection, and pointless or redundant illocutionary acts are not for that reason impaired as illocutionary acts in any way, nor are attitudes obviously inappropriate when their intended targets are out of reach. Rather, it is a conceptual or pragmatic difficulty we face of simultaneously taking two views towards the agent that conflict with each other.

It is a condition of satisfaction of blame, and probably of illocutionary acts generally that are directed towards another, that the speaker takes the addressee as competent to perform the so-called 'illocutionary uptake,' where in the case of blame this uptake is firstly to recognize that a legitimate moral demand to give a moral account has been issued. Something has gone wrong if we try to blame someone who we know perfectly well does not recognize such a demand and lacks the capacity to respond to it appropriately. This suggests that it is a constitutive rule rather than a regulative rule that the speaker believes that the addressee has such a capacity, for otherwise the speaker is making the addressee the target of reactive attitudes while knowing those attitudes to be inappropriate, and this would mean he was being knowingly inconsistent. When addressees have the capacity for *moral address* an exchange of reasons is possible, on which basis the appropriateness of the reactive attitude is determined.

Addressees' moral competence is correlated in this way with their competence in a certain kind of dialogue employing moral reasons in a process of challenge and

justification, and it is their status as legitimate moves in this kind of dialogue that unifies the multifarious considerations contained under the loose category of "quality of will" associated with the *excuses*. Note that in saying this I am not saying that moral competence and linguistic competence or moral norms and linguistic norms are in any way reducible to one another. The normativity is moral because the demand to give an account is moral, the reasons being exchanged in the account are moral, and because the blame we finally arrive at when all reasons have been considered "looks back at" these reasons and this demand.

So whoever performs the speech act of blame, by taking his blame as having at least potentially a communicative function, must take whoever the blame is directed at as capable of being *morally addressed*, that is to say, as having the kind of competence this requires, and this is a condition of satisfaction of the speech act of blame, or – if there are scruples about calling it blame at this early stage – whatever speech act it is by which the antagonist externalizes the community's moral demand that the protagonist provide a justification.[4] It might turn out that such a person does not have this competence, in which case the blame has an inappropriate target and is unwarranted, just as it is appropriate to withdraw our resentment if the person we were resenting is, for example, mentally incompetent and consequently exempt. This is one way in which our warrant for blame can fail; the other way is if the addressee has an excuse.[5] The condition of satisfaction that we believe our blame is warranted or, equivalently, that we believe the protagonist is responsible, is thus cashed out as the belief that no excusing or exempting condition obtains. Since it is by expressing and externalizing our reactive attitude of resentment that we typically (and indirectly) perform the illocutionary act of blame, it is unsurprising for the appropriateness of our blame and our resentment to correspond.

When we express a reactive attitude like resentment, then, we morally address an agent, where this indicates our assessment of the agent, communicates to the

---

[4] Is this condition of satisfaction a constitutive or just a regulative condition of blame? I am inclined to treat it as a constitutive condition, but frankly, I don't think it matters much. The illocutionary force of blaming is a performative and requires that the world be a certain way for its performance. The important point is simply that it is inappropriate and that we would take it to be inappropriate if we knew; whether we characterize this inappropriateness as failing to perform an illocutionary act or more specifically as failing to felicitously perform an illocutionary act does not matter much.

[5] Again, we might ask whether, when we blame someone who has an excuse, this is not a genuine case of blaming at all or simply an infelicitous blaming. A case could be made for the latter; non-obtaining of an excusing condition could be a regulative condition of blaming. A blaming has been carried out because the agent has knowingly and intentionally done something wrong and the demand to provide an account is itself appropriate because of the agent's capacity for *moral address*, but because they have an excuse no harm to them seems to be warranted. If so, this might be an interesting difference between excuses and exemptions, but not one that is material to the purposes of this paper.

agent that she has a case to answer and demand that she answer it. In a sense, it shows more respect towards the agent to treat him in this way than if one were to take an objective attitude towards him, for such an attitude does not recognize the agent as a fully-fledged moral agent but only as a kind of idiot to be managed instrumentally. Perhaps we might still call this objective attitude "blame", and it need not be assumed that by taking this attitude we need to change anything in our blaming practices, but this is not blame as we ordinarily think of it when we engage in these practices, as already said, and blame as we ordinarily think of it is not (while the objective attitude is) warranted solely by its social consequences.

McKenna's basic program is to provide an account of overt, directed blame on analogy with a conversation. In the first part of the next section I will give McKenna's account of this conversation he calls a *moral responsibility exchange*. In the second part I will describe the *critical discussion* model. In the third part I will compare the two. One obvious difference between them is that the critical discussion is a dialectical model and as such uses commitments rather than attitudes. It is an underlying reason for the comparison, then, to probe the possibility of 'dialectifying' the *moral responsibility exchange* and, by extension, the possibility of a theory of moral responsibility in which reactive attitudes have been replaced by what I call *reactive commitments*.

The possibility of 'dialectifying' the *moral responsibility exchange* does not stand or fall depending on its resemblance to a *critical discussion*, but if there is a strong resemblance then there is good reason to suppose that it can be done, that the exchange of moral reasons can be dialectified and thereby we can bring in the crucial notion of commitments, and furthermore we can reasonably expect the exchange of moral reasons to be regulated by the same rules as the *critical discussion*. For example, if a wrong-doer has the competence to defend himself but is never given the opportunity (i.e., the freedom rule is not observed), then this undermines our moral authority to blame him, and this seems to accurately model many of our moral and legal practices. This paper is a kind of feasibility study for this idea, though by the end I will be linking responsibility attributions more closely to *critical discussion* than a mere feasibility study suggests.

My hope is that this theory should have resources that the ordinary reactive attitudes theory does not in dealing with cases where there is no relationship with the one being blamed or emotion directed towards him. In this theory to say that an agent is appropriately held responsible will amount to saying that the agent has no winning

strategy in the dialogue and thus the exchange will conclude with a commitment to blame the agent, and this commitment will exist whether or not anybody actually blames or not, or whether anybody actually has a reactive attitude or not. If the agent does have a winning strategy in the dialogue then she is not responsible, either because the agent has a moral justification for what she did that is acceptable to her dialogue partners, or because it can be shown – again, through reasons that are acceptable to her dialogue partners – that the agent lacks the capacity for *moral address*. I am not proposing that having and not having a winning strategy are philosophical analyses of the concept of responsibility; rather, the idea is that our moral reasoning – or if not conscious reasoning then the recurrent patterns in the network of reactive attitudes and their modifications (for these are sensitive to and supposedly track moral reasons) – seems to take the form of a dialogue, and that dialectical notions (especially that of a *commitment*) can be usefully brought to bear on the problem of responsibility. Dialectified, McKenna's *moral responsibility exchange* becomes a decision procedure for deciding questions of moral responsibility, and the dialectical rules that tell us what moves are allowable at any particular point in the dialogue effectively map out the domain of possibilities for an agent's morally justifying whatever she finds herself under an obligation to justify.

Relationships and reactive attitudes should appear in the theory, but not to the extent that they become necessary conditions for holding an agent responsible or speech-act conditions for the illocutionary act of blame. It will also explain how *not* blaming can also be something for which someone can be held responsible. Beyond a few passing comments, this paper will not defend such a theory.

3. COMPARISON

  a) The *moral responsibility exchange*

There are three stages in the moral responsibility exchange: *moral contribution*, *moral address*,[6] and *moral account*.

        i)   *Moral contribution*

The moral contribution is the action to be assessed. There are two potential problems here that McKenna anticipates and responds to.

---

[6] Confusingly, *moral address* is called in some places the *moral demand*, e.g., McKenna (2012: 143 & 198).

R I A

The first concerns how an action can have the kind of content that it makes sense to say makes a contribution to a conversation. On this point McKenna (2012: 4) says:

> [I]nsofar as an agent's action is indicative of the moral quality of her will, it can be understood as itself an analogue to something like the initiation of a conversation by a competent speaker whose speech act instigates a conversational exchange… an agent's actions can themselves bear a species of meaning of interpretive significance – agent meaning – that is analogous to, but of course is not the same as, the Gricean notion of speaker meaning as distinct from sentence meaning.

The second problem is that the agent quite explicitly does not want or intend to communicate anything publicly by her action, on the supposition that she wants to "get away with" an act whose blameworthiness would at best be in question (McKenna, 2012: 93). This is a disanalogy with speech acts which typically at least involve communicative intentions. McKenna argues that this disanalogy does not threaten the usefulness of the analogy since we can still say that the agent means something by her act (the *agent meaning*), that it reveals something about her quality of will that perhaps she would rather remain hidden: "she thus acts with the understanding that her conduct is always a potential moment in the initiating stage of an analog to a conversation about the moral significance of her action" (McKenna, 2012: 98).

ii) *Moral address*

In the *moral address* stage

> those holding morally responsible respond to the blameworthy person in light of the meaning of her action, where her so acting is taken to initiate a conversation. This conversational response of holding morally responsible functions as a way of making moral demands, calling attention to expectations failed, and of course, expressing one's disappointment, anger, or frustration. Often there is an implicit demand for an account of some kind, an explanation, an apology, an acknowledgment of being in the wrong, an excuse, or perhaps a justification. (McKenna, 2012: 138)

There are two things that should be clarified here.

The first is that the *moral demand* being made is for the agent to give some account of their action rather than a demand not to perform the action, which would obviously have come too late to do any good, the *moral contribution* having already been made. Obviously, this demand depends on there being a demand/expectation not to perform the action and refers or "calls attention" to (perhaps by some kind of implicit anaphor) or expresses this originative demand in some way, but it does not *make* the

originative demand;[7] it is this originative demand that the *moral contribution* is deemed to have violated and that the backward-looking feature of the blame looks back at. It is through these stages of the conversation that this feature of the blame is publicly manifested, that is to say it is because the *moral* (not originative) *demand* is a response to the *moral contribution*, and it follows already from the *moral contribution* that the originative demand has actually been violated, the action having been given an *agent meaning* that defines it as something violating this demand. Existence of the violation is thus not in question. The moral reasons to be given in account do not, then, deny the violation, but argue instead that the violation was in fact permissible. Any argumentation about whether someone has done what they are accused of doing or about what *agent meaning* should be ascribed to the action is prior to the *moral responsibility exchange*. This is so for Strawson too, who only considers when reactive attitudes are inappropriate under the assumption that the agent did whatever she is said to have done. I will return to this later, but in the meantime perhaps an example might help.

You tread on my foot, and I express resentment towards you. You notice and say "Sorry! It was an accident!" The speech act of apologizing acknowledges that a moral demand has been violated, so we agree about this. By saying that it was an accident you are saying that by treading on my foot there was no ill-will or indifference involved, and this is reflected in the *agent meaning* of your act. If I agree with you about this as well, then I withdraw my reactive attitude. On this *agent meaning* no *moral contribution* has actually been made, and by the same token there is no confrontation between different attitudes towards a viewpoint, since we are agreed that a demand was violated but not on purpose. This is then prior to the *moral responsibility exchange*, and indeed such an exchange would itself be inappropriate in this context (though I am not sure this is McKenna's own position, for reasons that will appear later). But if you did it on purpose, then the *agent meaning* reflects this and a *moral contribution* has been made and there is a confrontation over the permissibility of this act: I am warranted in demanding that you give some kind of *moral account*, namely an excuse or an exemption, and the account you give will be linked by conversational mechanisms through the previous stages of the *moral responsibility exchange* back to

---

[7] This responds to the worry of Macnamara (2012) that it makes little sense to make a demand that, by definition, has already been violated. Nor does it make much sense to say that you are making the demand not to perform that action in future.

I find it curious that McKenna qualifies his claim as saying that this is a demand that is often made, and then only implicitly. On the contrary, I think that it is always made, and is made explicitly, at least in overt directed blame. Perhaps the qualification is there in order to cover cases where blame is not overt or directed.

the *moral contribution* and, implicitly, to the demand that is violated. I will show later that some kind of excuses seem to occupy a grey area where, although the violation may have been purposeful, no negative quality of will was manifested. In these cases too I do not think a *moral contribution* should be considered as having been made.

The second, perhaps more curious and problematic feature is that holding morally responsible, i.e., blaming, is located at this stage, which is to say, before the agent has had an opportunity to explain himself in the *moral account* stage. This seems problematic, as pointed out in Pereboom (2012: 193), since it seems to presume guilt and request the agent to prove innocence. Blaming an agent before they have given an account, and before or at the same time as they have even been asked to give an account, is morally in error, Pereboom says. Perhaps this is so when there is a reactive attitude but no *moral contribution*, as in some of the cases above when there is momentary resentment before the apology has been issued, but I will argue later that generally this is not an error.

Interestingly, McKenna's way of putting it seems to follow Strawson's (2008), who tends to talk in terms of revising attitudes that you have already formed, suggesting, perhaps, that the reactive attitude is an emotional reaction that is only afterwards brought under the control of reason. McKenna is also helped here by distinguishing blame from punishment; blame is warranted in *moral address*, but punishment is not warranted until everything has been considered — it has conditions that mere blaming does not (McKenna, 2012: 91 & 144). In contrast, Wallace (1994) seems to put the judgment first and the emotion second. Later on, I will give reasons why I think McKenna is right about this, although I will also partially agree with Wallace that being responsible does not depend on anybody actually having a reactive attitude, the dialectical notion of a commitment performing much the same duties that Wallace gives to the judgment.

iii)   *Moral account*

It is in the third and final stage of moral account that the agent gets to defend herself "either by appeal to some excusing or justifying consideration or instead by way of an acknowledgment of a wrong done, perhaps an apology offered" (McKenna, 2012: 89). If the agent fails a capacity condition, then she will likely be practically unable to give an exempting reason, but McKenna (2012: 89 ff. 8) says that one could be given on her behalf. This conversation may continue with acts of forgiveness or penitence. "Each subsequent stage in the conversation offers the interlocutor reasons to consider in a

new light either the moral quality of the agent . . . the relationship . . . or, depending on the nature of the account offered, the initiating conduct" says McKenna (2012: 89). Thus, the norms governing the conversation and the appropriate forms of sanction seem to be set by the *moral contribution* and by the particularized relation between the interlocutors. Only a wife may be entitled to blame her husband, or sanction him in particular ways, for certain actions, but there are other actions for which anybody at all might be considered to be so entitled (McKenna, 2012: 195-200).[8]

Responsibility is defined negatively as the non-obtaining of an excusing or exempting condition (and, perhaps we should add, the non-violation of the conversation's regulative rules). Thus, we can start to see in McKenna's *moral account* the germ of the idea that an agent is appropriately held responsible if he cannot give a satisfactory *moral account*, which is to say that he cannot 'win' the dialogical exchange of reasons.

b) The *critical discussion*

There are four stages in the critical discussion: the *confrontation stage*, the *opening stage*, the *argumentation stage*, and the *concluding stage*.

i)   The *confrontation stage*

In the *confrontation stage* "a difference of opinion presents itself through an opposition between a standpoint and nonacceptance of this standpoint. . . . . If there is no confrontation of views, then there is no need for critical discussion" (van Eemeren, Grootendorst, and Henkemans et al., 1996: 281-82).

ii)   The *opening stage*

In the *opening stage*

> the protagonist and the antagonist in the dispute are identified and their initial commitments – substantive, procedural, and otherwise. The protagonist undertakes the obligation to defend the standpoint at issue, while the antagonist assumes the obligation to respond critically to the standpoint and the protagonist's defense. In argumentative discourse, this stage corresponds with the phase where the parties manifest themselves as such and determine whether there is sufficient common ground (shared background knowledge, values, rules) for a fruitful exchange of views. It only makes sense to undertake an attempt to eliminate a difference of opinion by means of argumentation if such a starting point can be established. (van Eemeren, Grootendorst, and Henkemans et al., 1996: 281-82)

---

[8] However, Bell (2012) denies the intuition here; there may be something wrong with our blaming without it being the case that blame is inappropriate. Indeed, there seems to me a possibility here for there to be some kind of equivocation between distinct kinds of appropriateness.

iii) The *argumentation stage*

In the *argumentation stage*

> the party that acts as the protagonist methodically defends the standpoint at issue against critical responses of the antagonist. If the antagonist is not yet wholly convinced of all or part of the protagonist's argumentation he or she elicits new argumentation from the protagonist, and so on. . . . In argumentative discourse, the argumentation stage corresponds with the phase in which one party adduces arguments in order to overcome the other party's doubts about the standpoint, and the other party reacts to those arguments. Adducing argumentation, and judging its merits, is crucial to resolving a difference of opinion. (van Eemeren, Grootendorst, and Henkemans et al., 1996: 281-82)

iv) The *concluding stage*

In the *concluding stage*

> the protagonist of a standpoint and the antagonist determine whether the protagonist's standpoint has been successfully defended against the critical responses of the antagonist. If the protagonist's standpoint needs to be withdrawn, the dispute has been resolved in favour of the antagonist; if the antagonist's doubts have been retracted, it has been resolved in favour of the protagonist. In argumentative discourse, the concluding stage corresponds with the phase in which the parties draw conclusions about the result of the attempt to resolve the difference of opinion. If the parties do not agree on the outcome of the resolution, then the critical discussion has not led to a resolution of the difference of opinion. (van Eemeren, Grootendorst, and Henkemans et al., 1996: 281-82)

c) Comparison of the *moral responsibility exchange* with the *critical discussion*

The first obvious disanalogy between the moral responsibility exchange and the critical discussion is that critical discussions have four stages and moral responsibility exchanges only three. Clearly, the *moral account* maps pretty naturally onto the *argumentation stage*, for this is where reasons are exchanged. One would hope that the *moral contribution* could be mapped onto the *confrontation stage* and the *moral address* onto the *opening stage*, albeit with a little shoe-horning. What, then, of the *concluding stage*? Does this have no corollary in the *moral responsibility exchange*?

I will argue that the *concluding stage* is not properly part of the critical discussion. Thus, both the critical discussion and the *moral responsibility exchange* have three stages, with the first stage of the critical discussion analogous to the first stage of the *moral responsibility exchange*, and so on as mentioned above. The *concluding stage* is analogous to something that McKenna discusses, namely a punishment stage. But McKenna is clear in placing punishment beyond the stage of *moral account*; it is not part of the *moral responsibility exchange* as such. If, however,

R I A

the concluding stage is, as orthodox pragma-dialecticians hold, part of the critical discussion, then the *moral responsibility exchange* can be revised so that the punishment stage is counted as an additional stage. In this scenario, both the critical discussion and the *moral responsibility exchange* have four stages.

Why, contrary to orthodox pragma-dialecticians, do I say that the concluding stage is not a part of the critical discussion? What seems to be intended in this stage as described above is that the actual parties to the critical discussion respond appropriately to the dialectical situation by modifying their psychological attitudes and regulating their behaviour as prescribed or proscribed on the basis of whether the standpoint is conclusively defended or defeated; to put it another way, a conclusively defended standpoint becomes a premise for practical reasoning such that acting contrary to it is by definition criticizable. Once they have agreed to such a resolution, a party must abide by it and can be criticized if he does not, even if he still does not believe the standpoint. Also, if they do not actually agree that a resolution has been found then the critical discussion may either be continued aimlessly without the possibility of ever being closed, or it may be simply abandoned. For the standpoint to be established, not only as the conclusion of the preceding dialogue but as a constraint on ensuing behaviour, there seems to be needed an agreement that this is the resolution. Equally, if there is agreement that there is no such resolution, then it would seem that neither participant is criticizable by the other for behaving contrary to the standpoint, for the standpoint has been defeated (note that this does not mean that it has been shown to be false but only that it has not been shown to follow from the starting-points) and so there is no constraint on ensuing behaviour with regard to it. There must be agreement either way (that there is a resolution or that there is not) in order for certain normative standards on participants' behaviour to be in force and their enforcing of those standards to be justified. Where there is no such agreement then it might be thought that neither participant can be criticized by the other, or at least not justifiably. I will challenge this, however; lack of agreement does not mean that no normative standards apply, and I will argue that when one participant rightly believes that there is a resolution then a participant that will not agree is criticizable.

There are two things that should be noticed here, and which are my reasons for separating off the *concluding stage* from the critical discussion.

The first is that it evinces a slide from a commitment-based model to something like a belief-desire-intention model, for what is being talked about here is not in terms of dialectical concepts of protagonist, antagonist, and commitment, but the attitudes and

behaviour of flesh-and-blood arguers taking those dialectical roles; as such, it does not seem to be something that can be modelled dialectically, or properly be part of the dialectical process that the critical discussion delineates.

The second is that, even if it is true in practice that arguers may continue to disagree unless they actually agree that a resolution has been found, this does not alter the fact whether a resolution has been found, for this is determined by whether there is or is not a particular commitment, viz., the standpoint. There is a sense in which there is a resolution, even if nobody knows that there is: the critical discussion is a dialectical model that works in terms of commitments which, along with the rules of critical discussion, define the dialectical situation, and once all reasons have been offered and subjected to criticism, once all lines of argumentation have been exhausted, the dialectical situation is settled and we can say with complete confidence whether the protagonist has a winning strategy, i.e., whether the standpoint should be allowed to stand,[9] and while this may be a concluding *point* of the dialectical process it makes little sense to call this a concluding *stage*, or to say that there can be any such stage in the dialectical process as such beyond this point. If some party does not agree then one party or the other has made a mistake and is being criticizably (though perhaps not consciously) unreasonable because refusing to accept a shared commitment.[10] In short, it should not be a part of the *dialectical* model that "If the parties do not agree on the outcome of the resolution, then the critical discussion has not led to a resolution of the difference of opinion." This does not seem to be a dialectical rule, strictly speaking.

---

[9] The notion of a winning strategy comes from Barth and Krabbe (1982: 315) rather than pragma-dialectics. This marks one difference between my account and orthodox pragma-dialectics: pragma-dialectics only requires the protagonist to respond to questions that have actually been asked, and so her standpoint is defended – she "wins" – if the antagonist does not ask any questions, though pragma-dialecticians would stop short of saying that a standpoint that has been insufficiently critically tested is reasonable. But having a winning strategy does not depend on what questions a person can think to ask but covers the whole space of dialectical possibilities, and it is the existence of a winning strategy that counts rather than a flesh-and-blood arguer actually taking that strategy.

[10] Accepting a commitment, however, does not necessarily mean believing it to be true; it means only taking it to be true qua member of the group comprising both parties whose joint commitment this is. This is more fully explained in my paper "Being reasonable" (2015) and is an amendment to orthodox pragma-dialectics. My basic premise is that dialogue, even adversarial dialogue, is a co-operative activity with a certain goal, and that all parties to the dialogue form a group identified by their shared commitments agreed to in the opening stage. The moves by protagonist and antagonist should then unfold, by making explicit the consequences of those shared commitments (what has been called the dark-side commitments), what the group as such should accept. It may be that some party, on discovering some such commitment, would prefer to retract one of the shared commitments. This party may request such a retraction, but may not (because then he would be abandoning a goal that he had previously agreed to co-operate in achieving) demand a retraction, nor is the other party obliged to grant such a retraction. If the retraction is not granted, the party is nonetheless obliged to accept what he may well not believe qua individual, and this, I argued, is a moral obligation that if broken invites moral sanction, as well as being criticizably unreasonable. This is so both for the new commitment and the commitment whose retraction was requested. These commitments must still both be accepted, though they will not in fact be believed.

If, however, they do not agree on the outcome of the resolution, does this mean that the arguers can go on arguing indefinitely, either in good faith or in bad? They could, but not without breaking some rule of the critical discussion, such as the one that forbids repeating the same argument. It should be the rules that define when the critical discussion is over and not the arguers. The critical discussion as dialectical process is closed when its rules say that it is closed, and cannot be continued, or denied to be closed, or agreed to be unresolved, without violating one of those rules,[11] and whoever breaks the rules is a candidate for criticism.

Similarly, punishment is not part of the *moral responsibility exchange*. In a case where the critical discussion concerns a case of moral responsibility, the behaviour of arguers that seems the focus of the concluding stage takes the form of considering what punishment or coercion may be appropriate and carrying it out, that is to say, it concerns regulating behaviour subsequent to and with respect to the closure of the *moral responsibility exchange*. That is to say, this *concluding stage* where arguer's behaviour is prescribed or proscribed on the basis of whether the standpoint is conclusively defended or defeated, is analogous to the stage immediately after the *moral responsibility exchange* where questions of punishment and sanction are to be decided and carried out.

The result is that we can either treat both as stages in their respective models, or as stages beyond what the models are attempting to delineate.

This removes one obstacle to identifying *moral account* with the *argumentation stage*. It also seems to deal with one possible problem that was mentioned above, for recall that McKenna's account was primarily an account of *directed* blame and its relevance to cases of blame where there was no-one present to be blamed (and this follows also for cases of *private* blame where it was not even made manifest or overt)

---

[11] It is true that the rules refer to dialectical moves such as defending or attacking a standpoint, but it should be remembered that although these moves are realized in real argumentation by arguers performing certain speech acts the rules themselves do not refer to flesh-and-blood arguers or argumentation. A protagonist, for instance, is not a person but a dialectical role. The effect of my views on the concluding stage is to ameliorate the role of the intersubjective testing procedures etc. by which parties to the argumentation may agree that there is a resolution. Although it is perfectly true that this determines in part what the parties may do and what behaviour is criticizable, it does not affect the dialectical result: a standpoint is either defended or it is not, and there is a resolution one way or the other. If the standpoint is conclusively defended but the antagonist refuses to accept the standpoint because he refuses to agree that all the inferential steps comply with argumentation schemes that were agreed to in the opening stage (although they do), then I am inclined to think that the protagonist is well within his rights to criticize the antagonist for being unreasonable. There will be no real resolution of the disagreement, but there will nonetheless be an outcome of the critical discussion. On the other hand, if they *agree* that the standpoint is *not* conclusively defended then it does not seem reasonable for either to criticize the other for acting inconsistently with it, even in the event of their both being wrong and the standpoint's being conclusively defended.

was questioned. However, blame *qua* commitment does not have this problem, because it is quite indifferent to whether or not blame is effective, is manifested, or even whether there is an attitude corresponding to it in the one who would blame.

The commitment to blame is generated as a *dark-side commitment* by the opening stage and is there in the dialectical situation when the critical discussion concludes, and will be so even if nobody becomes aware of it; the moral community as a whole will have a dialectical obligation and a moral responsibility to blame. That this has moral normativity and not just dialectical normativity is guaranteed by the moral normativity embodied in the *moral confrontation* that initiated the exchange and to which all later dialectical commitments can be tracked back; in particular, as I have already said, it is the original moral demand that was violated that the blame looks back at, and the dialectical moves carried out, which it should not be forgotten are dialectifications of exchanges of *moral* reasons, are the mechanism by which this looking back occurs in dialogue. It is not that dialectical norms replace moral norms; a reactive commitment, though it is a commitment in dialogue, still has a moral normativity.

However, there are possible disanalogies between *moral account* and the argumentation stage that must be considered further.

i)   Comparing the *moral account* and the *argumentation stage*

It is interesting with regards to the conversational theory that Strawson himself described the excuses and exemptions functionally in terms of the kinds of things one would say in dialogue. Exemptions obtain when the agent (or someone on the agent's behalf) could say in response "He wasn't himself," "He has been under very great strain recently," "He was acting under post-hypnotic suggestion," "He's only a child," "He's a hopeless schizophrenic," "His mind has been systematically perverted," or "That's purely compulsive behaviour on his part." Excuses obtain when the agent could say in response "I didn't mean to," "I hadn't realized," "I didn't know," along with things like "I couldn't help it" when backed by "I was pushed," "I had to do it," "It was the only way," or "They gave me no choice" (Strawson, 2008: 7-8).[12] Strawson has here loosely

---

[12] Strawson puts these excuses in the third person, which implies that it is not necessarily the accused who is answering. Yet it is the accused to whom the demand for an account has been addressed and who seems obliged to answer, Watson and McKenna seem to say. Is this a problem? McKenna, it has been noted already, allows for exemptions to be given as reasons on behalf of a third person, and I see no principled reason to restrict this to exemptions. If the moral community can say on behalf of the accused that, e.g., he didn't realize what he was doing, they may do so, and in doing so they take for that 'turn' in the dialogue the role of protagonist. Although it is the *moral contribution* of the agent that initiates the dialogue and identifies the agent as a protagonist defending the standpoint that her action was permissible, this does not mean that she and only she may take the role of protagonist.

separated two groups of excuses but not clearly distinguished between them, characterizing both as manifesting the quality of will. I think there is a difference between them.

I think we can see the difference more clearly if we compare it against legal procedures. A defending attorney has four basic strategies: to plead not guilty, and argue that the defendant simply did not do what she is accused of doing; to plead guilty, apologize, and throw the defendant on the mercy of the court; to plead guilty, but argue that there was no intent (mens *rea*); or, to argue not guilty on grounds of diminished responsibility. Strawson is not interested in the kinds of cases where the accused simply did not perform the action, but perhaps he should be, because some excuses come to very nearly the same thing.

It needs to be noted in this regard that the action at issue is not simply a bodily movement; no moral issues or difference of opinion arise over the fact that I moved my finger, and no *moral contribution* is made when the action is given this *agent meaning*. However, if this finger movement results in someone being shot and killed then the agent can be accused of performing the action of murdering someone. On the other hand, perhaps the agent believed the gun was unloaded and it was just a kind of accident. In this case, it is inaccurate to call this action "murder." A reason of this kind, then, denies that the agent performed the action of murder, though conceding the action of moving his finger. In this way it is in fact quite similar to the more fundamental claim expressed by the words "I didn't do it," and there probably are cases where we do express it this way. "I didn't mean to," "I hadn't realized," "I didn't know," come very close to this, denying *mens rea*, which is to say, denying that the action had the *agent meaning* that it had been presumptively interpreted as having, and recall from our original program statement that "those holding morally responsible respond to the blameworthy person in light of the meaning of her action" rather than to the action as such. These kinds of excuses are not, then, argumentative in the same way as the others but are clarifications or correctives of how the action is to be interpreted; it is prior to the moral responsibility exchange and the critical discussion.[13] If the interpreter agrees with this interpretation, any further exchange of reasons is pre-empted. These

---

[13] We have already seen that McKenna says that in the unfolding of the conversation we may be brought to re-assess the initiating conduct. I don't think he would agree with me that this makes it in any way prior to the moral responsibility exchange, however. My interests are to reconstruct the exchange in a way appropriate for its evaluation, and this means that any ambiguities concerning the standpoint must be resolved at the analysis stage and prior to exchanging reasons for and against the standpoint, which in the end is what we need to evaluate in order to hold someone responsible. They may not be prior with respect to where they occur in the actual conversation.

excuses work similarly to "Sorry! It was an accident!"

Apart from these kinds of responses and excuses that work like correctives of the interpretation, it is not (*contra* Pereboom) unreasonable to respond with a reactive attitude before the agent has given any *moral account*; in essence, guilt has been admitted, and the defending attorney is resting his case on showing one kind or other of diminished responsibility. Exemptions would obviously fall into this category: incapacity to know right from wrong, for instance, being clearly a case of incapacity warranting a verdict of diminished responsibility. But the excuse "I couldn't help it" may also be another kind of diminished responsibility, such as instances where the accused was acting under instructions from criminals who held his family hostage. Let us reconsider your stepping on my foot and suppose that instead of "Sorry! It was an accident!" you respond to my resentment with "They made me do it!" You here agree with me that a moral demand was violated, on purpose, and that the action as such seemed to be a manifestation of ill-will; although perhaps you felt no malice towards me, you did not care sufficiently about the moral demand to prevent yourself from violating it. If "They made me do it!" is cashed out as "They would hurt my family if I don't" then the agent does not seem to be blameworthy (though the agent may still be responsible in at least some senses of this word) and her violation seems permissible; thus, the standpoint that the act with this *agent meaning* was permissible in the agent's actual circumstances would be defended and there would be no commitment to blame. On the other hand, if "They made me do it!" is cashed out as "My friends dared me to!" this does not seem to be the kind of reason that makes violation of the demand permissible. Between these extremes there are a great many intermediate cases, and I think that what makes "They made me do it!" assertible as an excusing reason in a *moral account* probably varies from one moral community to another. Thus, there is a certain amount of relativism present in this view of moral responsibility. Excuses and exemptions are defined functionally – and on the view presented here, dialogically – not substantively, and hence they can vary substantively from community to community and from person to person within the functional parameters.

There is also another way reasons might operate. It may be agreed by all involved that the agent is blameworthy and that the moral community is aware of its dialectical obligation to blame the agent, and yet no particular member of the moral community actually blames the agent. There are a number of reasons why this may be so: it could be that the would-be blamer, imagining himself in the agent's position, decides that he would have done the same, that under the same pressures he would

have acted as the agent acted, or; it could be that the would-be blamer does not believe themselves to be in the right particularized relationship to the agent, such as being the wife of the abusive spouse, or to have enough of or the right kind of authority, or; it could be, quite simply, that in the absence of experiencing a reactive attitude he does not feel entitled to cause harm. These are the kind of reasons a would-be blamer might rightly or wrongly take to defeat their own personal warrant for blame, without defeating the moral community's warrant. This is where I believe being responsible and being held responsible can come apart: an agent is responsible if the community's warrant is undefeated, though nobody in the moral community holds the agent responsible either because they think that they themselves have an excuse or other responsibility-undermining reason for not blaming (meaning that they are not blameworthy for not blaming, to be decided by another moral responsibility exchange) or because those with the authority to hold the agent responsible think (rightly or wrongly) that this authority is undermined.

Something like the latter happens in the following case: it is another kind of response that I would also characterize as a corrective, but of a different kind that does not dispute the agent meaning of the action but the force of the blame. This is the kind of response expressed by the words "What right do you have to judge me?" when backed by things like "You are just as bad yourself!", "You would have done the same," "You have done the same yourself," and, more tersely, "You hypocrite!" This seems to generate conflicting intuitions, on one side such a response as "That's irrelevant! Do as I say and not as I do!" to "Yes, you're right; it is self-righteous of me to resent you for doing what I have done/would do myself."

It is interesting that neither interlocutor is disputing the facts of the case — they may well agree that the agent is blameworthy. Where, then, is the defect in blaming him? How are we to understand the agent's complaint? Remember that there are different aspects of blame – a forward-looking aspect and a backward-looking aspect – and it was the backward-looking aspect that was linked to the reactive attitudes and that was taken to give blame a distinctively moral force. But there are also blamings or at least sanctions that are not taken to have this kind of moral force, such as when their warrant is given entirely by forward-looking considerations, e.g., social regulation. The same sanction can be seen as the externalization of either a reactive attitude or an objective attitude, and may be warranted as the latter but not the former. Thus, "Who are you to judge me" could be glossed as "I agree that I have done something blameworthy and perhaps I should be incarcerated in order to protect society. Also, I

do not dispute your legal authority to have me incarcerated. But I do not accept your moral condemnation or your moral authority." By this she asks for an objective attitude and an instrumentally-warranted sanction; it is only to these that the other is entitled, according to the complaint.[14] Perhaps she might also allow that the other is entitled to moral evaluation, but not to blame with the kind of force that we ordinarily give it, because to be entitled to that force the blamer must be consistent — at least, so goes her intuition. What is more, from what I said above it seems that the authority himself might agree, and (rightly or wrongly) foreswear blame in this sense because believing it is inappropriate for him personally to blame, though appropriate for the moral community as such.

But this has led us back in a circle, for we have not answered the question of whether a blamer must be consistent and, if so, why. Wallace's answer is, briefly, that when we blame hypocritically we are implicitly taking ourselves to be of a different moral standing to the one we blame, and by doing so fail to respect our equality as persons (Wallace 2010). I find this less than obvious and, in fact, I do not think that a blamer must be consistent. In fact, I will argue later on that when I blame an agent for an action I implicitly blame all relevantly similar agents performing relevantly similar actions. If I myself have performed such actions then I fall within the scope of my own blame, and I may indeed realize this and deeply regret my past transgressions and see them for what they were and see the excuses and justifications I may have appealed to at the time to be the self-deceptions they were. The 'inconsistency' involved here seems to be nothing more or less than conscience. Thus, there is no moral fault in hypocritical blame — it is often the first step on the road to moral realization and non-hypocritical blame.

With these kinds of excuses it is not denied that a morally bad action was intentionally performed by the agent (and that thereby the agent has made a particular kind of *moral contribution*). With the exception of the maker of the hypocrisy complaint,

---

[14] What is possibly more likely is that the one making this objection does not make this distinction and consequently mistakenly believes the objection should shield him from any punitive measures at all. To give an example: in an episode of British comedy series "Porridge" career criminal Fletch has to share a cell with the judge who convicted him, who has now also been convicted of a crime. Fletch feels that an injustice has been done, that even though he was guilty of the crime he was convicted of the judge had "no right." To this complaint it seems quite right for the judge to reply that his right was "The rights that were properly invested in me in my position as a judge." The judge's crimes are irrelevant to the verdict he quite correctly handed down to Fletch and the punitive action taken by the moral community through the office of the judge is warranted. If, however, the judge expressed personal indignation towards Fletch then perhaps he might be justly accused of hypocrisy. However, although this indignation may manifest a reprehensible quality of will that the judge may legitimately be held to account for (such as a superiority complex that "the rules apply to other people but not to me") it is not obvious to me that the indignation is itself unwarranted.

the intuition is that the agent is not morally blameworthy in these situations, and Strawson would say that this intuition is because the agent is not demonstrating ill-will towards us.[15] However, there is apparent ill-will, and so initially it at least seems reasonable to hold this agent responsible (though not to punish him) prior to the *moral account*. It is these kinds of considerations that *moral account* will typically raise, then, and on this understanding *moral account* corresponds closely to the *argumentation stage*.

Suppose lastly that you respond to my resentment at having my foot being trodden on with "I'm a hopeless schizophrenic," or, what is perhaps more likely, someone gives this on your behalf, or perhaps again when I look at you more closely I notice features that seem to disqualify you from full participation in interpersonal relationships. In other words, you lack the capacity for *moral address*, for a certain kind of competence. What happens now to the *moral address*? Clearly, it fails to perform its characteristic communicative function and, recognizing this, I should withdraw my reactive attitude and possibly substitute an objective attitude in its place. Here, though, a *moral contribution* has been made, for it is not the quality of will that is relevant to these cases (making no assumptions about whether you harbor malice or ill-will towards me or not), and a *moral address* has likewise been made, albeit one that cannot communicate and that for that reason should afterwards be withdrawn because inappropriate.

ii)   Comparing the *moral contribution* with the *confrontation stage*

There is a *moral contribution* when the *agent meaning* is such that a confrontation arises from it. As said above, such an *agent meaning* could not be simply that the agent has performed some bodily action, but that the agent has performed some action that violates a moral demand. The *agent meaning* is meant to indicate the agent's quality of will in performing this action. The agent may argue about what interpretation her acts should be given but cannot deny them an interpretation at all by saying that her action was never meant to be discovered. From the interpretation a commitment is formed. As McKenna says, this initiates the *moral responsibility exchange*, and likewise sets up the agent as the protagonist in the *confrontation stage* of the *critical discussion*.

What, exactly, is this commitment though? It is already, in a sense, agreed that the action can be accurately identified as one to which reactive attitudes are an apt

---

[15] Note that there could be ill-will even in this situation, if, for instance, an agent was blackmailed into doing something that the agent wanted to do anyway (and perhaps endorsed that want). Strawson seems to be right that our intuitions track the quality of will.

response. The commitment – the standpoint that the protagonist must be taken as defending – is that the moral demand was one that it would not be blameworthy to violate under certain conditions that actually obtained, and the confrontation arises because the moral community is not committed to this standpoint. Diminished responsibility need not mean zero responsibility, however, and we need not suppose that every moral community would consider responsibility to be diminished by the same reasons and by those reasons to the same degree as another moral community. It seems to me a very real possibility that an agent may be responsible to more than one moral community and may appropriately be held responsible by one community but not the other. If we are to accept the Strawsonian idea that to be responsible is to be held responsible (both of these being triadic relations as a consequence), then we have the result that an agent is and is not responsible. I do not think this is a contradiction but (we will see in the next section) naturally emerges from the *opening stage* of the *critical discussion*, the *argumentation stage* being built upon the material starting-points agreed to in this particular dialogue and in that sense relative to this particular dialogue.

This is a non-mixed dispute, and it is the agent – the accused – who has all the dialectical obligations to give reasons. The blamer does not need to give reasons to blame (though he may need to give reasons to punish); blame is the default position because it is the appropriate initial response to the *moral contribution*. Resentment can be attributed at the *moral address* stage to the moral community on the grounds of the *agent meaning* alone, since for the act to have made a *moral contribution* it must be attributed an *agent meaning* such that it violates a moral demand, as already said. Resentment must only be withdrawn later if the protagonist gives a successful *moral account*. The matter is different when it comes to punishment or sanction, for then a whole raft of new reasons comes into consideration, and the moral community may then have to argue that it is justified in coercing an agent or depriving them of their liberty, and the authority needed for this is much higher than that needed for overt blame.

iii)   Comparing the *moral address* with the *opening stage*

The *moral address* for McKenna is simply the holding responsible of the agent. There is a disanalogy between the way this relates to the *moral contribution* and the way the *opening stage* relates to the *confrontation stage*, because what the agent is held responsible for is the action itself rather than the standpoint and holding responsible is a response to the action rather than the standpoint. I am prepared to accept this disanalogy; I do not think it too important.

A more difficult problem is to find in the stage of *moral address* something like what happens in the *opening stage* of the *critical discussion* where all the premises and procedures that can be appealed to in the *argumentation stage* are decided on. On one level, these are already decided on for Strawson: they are the considerations and rules that already regulate our reactive attitudes and are given to us as part of our human condition and do not need any further justification. In this sense, protagonist and antagonist have no real option but to 'agree' to these.

Although it might seem a bit misleading to think of this as agreement, there is another sense in which this fits what Strawson says very well, for it is essential to Strawson's view that any modification of our reactive attitudes are due to application of the internal rules regulating them and cannot be affected by any consideration applied as an external criterion, and this is precisely what a *critical discussion* says when it says that you can only appeal in the *argumentation stage* to starting-points and inference schemes that were agreed to in the *opening stage*. If some metaphysical thesis like determinism was true it could not *qua* external criterion bring us to revise our reactive attitudes but only by realizing an internal criterion, that is to say, by being the kind of thing because of which it could be said, e.g., "They made me do it," "He couldn't help it," etc. But if the internal criteria are defined by these kinds of functional criteria, it seems possible that different communities may take different considerations as satisfying these criteria, and to different extents.

What I am trying to find in *moral address* is something loosely corresponding to agreement to starting-points. I think we can do this if we emphasize that the antagonist is, first and foremost, a moral community as such, rather than the particular person injured. It is then part of what it is to be that moral community that certain expectations, demands and practices are in place. Being *responsible to* the moral community, then, is to be committed to those expectations, demands and practices. This does not necessarily mean that the agent believes those demands are the right ones and the practices to be warranted, but that she accepts them *qua* member of the community.[16]

---

[16] It would lead me too far from my current subject to discuss the kind of collective intentionality involved here and the conditions under which it is realised. Suffice it to say that it is of a kind such that the group can legitimately make moral demands of its subjects, and subjects have moral obligations towards the group to comply with those demands. In particular, the group can legitimately morally demand that the subject provide a justification of actions that transgress the norms of the group, and this demand is binding on the subject such that she is obliged to respond to it and cannot escape by rejecting the group's authority to make that demand or by somehow opting out of the moral community. The conditions for membership in the moral community are weak and do not require, in particular, that the subject actually agree with the group's norms. If a moral community holds you responsible for a murder, for example, then saying that you, or even the culture that you grew up in, finds murders of this kind morally permissible and consequently that you disagree with the norm that says it is not, will not provide you with a free pass on

The resolution aimed for by the critical discussion is ultimately a joint decision by the community as a whole, including the agent unless the agent is one who cannot be engaged with in the dialogue (in other words, unless she is exempt and hence not a candidate for membership in the moral community). If she loses (has no winning strategy) in the *argumentation stage* then in some sense the agent must agree with and be committed to the standpoint, in this circumstance to being blamed. Agreement, then, is implied by being responsible to the community in question in the first place. There is no explicit agreement to starting-points, but this is typically the case for dialogues modelled as *critical discussions* as well.

This might seem to reduce the role of the victim, and deny in some way that the victim may have rights and privileges with regard to the action that perhaps the average community member does not. I do not think this follows. Of course, it is very likely the victim who will be prone to the reactive attitudes. Other members of the community may not have those attitudes, even if they know about the situation and agree that the action was morally blameworthy and deserving of a negative moral evaluation. What I have in mind is somewhat different, and it is this: the one injured, in reacting to this particular token of a violated demand, is implicitly reacting to all relevantly similar tokens, that is to say that he conceives his reaction to be an appropriate moral response, and in conceiving it to be such, the reactive attitude appears to him to be what any relevantly similar subject made to suffer a relevantly similar injury by a relevantly similar agent would be warranted in having, and if they would not be so warranted, then neither is it warranted in the particular case. It is simply part of the phenomenology of moral sentiments and the standards of appropriateness that we take to apply to them that they universalize in this way. It is a false dichotomy to ask whether the injured party is responding *qua* individual or *qua* member of the community. However, one may have an attitude or acceptance *qua* member of the community that one would not necessarily endorse *qua* individual. Nor am I saying that the individual's having the reactive attitude makes it true that the reactive attitude should be attributable to the community as a whole; how the attitudes that are attributable to the community supervene on those that are attributable to its members can be a complex affair that can itself be modelled dialectically as commitments generated by agreed-upon rules.[17] Lastly, even when a reactive attitude

---

the questionable grounds that in virtue of this disagreement you are not a member of the moral community and, being such, are not bound to conform with its expectations; whether cultural or personal differences over norms will provide some level of mitigation will be decided according to the substantive norms of the community and not by philosophical analysis. These are complications I have dealt with elsewhere.

[17] List and Pettit's (2006) premise-based procedures lend themselves easily to this kind of modelling. Their

can be attributed to the community, it still remains the fact that when the injured party blames, it is his own reactive attitude that he is expressing, in spite of the fact that the warrant for the blame comes from the norms of the moral community (as they are worked out in the *moral responsibility exchange*), as does its force when it has a force (i.e., when the injured party does have the authority to blame).

4. CONCLUSION

In this paper I have argued that the *critical discussion* is a good dialogue model for the *moral responsibility exchange*. In so far as it models the exchange of moral reasons, it is also a good model of moral reasoning, or at least reasoning about moral responsibility. It also explains some peculiar empirical results in moral reasoning, such as why we hold someone responsible for a negative foreseeable side-effect that they were aware of but were indifferent to but do not praise them when the side-effect is positive: I think this is plausibly because the agent is obliged to give and could not give an adequate justification when the side-effect is negative, that is to say, the agent would not have a winning strategy. Our reasoning follows the contours of this dialogue and our resulting judgment of their blameworthiness tracks the outcome of this dialogue, even if we are only imagining the dialogue. When there is such a dialogue, though, the model of a critical discussion seems to model this dialogue so that particular illocutionary acts of blame have the communicative function we should expect and have the force and warrant we implicitly take it have in our ordinary conception of blame, and moreover the rules of critical discussion seem to regulate the dialogue in a way that leads us to a successful resolution and that we seem to follow.

Pragma-dialectics is quite comfortable with the idea of an action being interpreted as the indirect performance of a speech act, even without the elaborate defence McKenna gives of *agent meaning*; it is an advantage of the dialectical

---

purpose is to make it so that the attitudes of the group are consistent, and there are cases where for this to be so the attitudes of the group must clash with some of the attitudes of the individuals comprising the group. These are so-called discursive dilemmas, and in some such dilemmas an attitude such as belief in a particular proposition must be attributed to the group even though no member of the group actually believes it, but because it is the logical consequence of other propositions that they have accepted. The amenability of this to dialectical modelling should be obvious, and reinforces the view that it is the group as a whole, i.e., including both the protagonist and the antagonist, that are *jointly* to come to a resolution and agree to or reject (but not necessarily believe or disbelieve, note) the standpoint.

Incidentally, this is why it does not matter that one subset of a community has moral beliefs that another subset does not. There may be conflicting beliefs, but not conflicting acceptances, for the procedures are designed precisely to avoid any such inconsistency at the group-level. The acceptances of the group are closed under entailment, and, in List and Pettit's view, it is pursuing closure at the group-level rather than the individual level that makes the group akin to an intentional subject to which it makes sense to attribute cognitive attitudes such as beliefs and also, I would argue, reactive attitudes like resentment.

commitment-based model that it does not depend on defending an analogy between actions and speech acts and between practices of blame and conversations that McKenna admits becomes stretched in places. Even if the blameworthy agent does not intend her action to be public knowledge there is still a commitment to the standpoint that her action, although morally bad, was morally permissible, and even if she is not there to be blamed at the close of the *moral responsibility exchange* this *moral contribution* warrants in the *moral address* stage a further commitment to hold the agent responsible and to demand a *moral account* of what the circumstances are that would in this case render the action permissible and warrant retraction of the commitment to blame (if given successfully using only the starting-points agreed to, which is to say, only those that would be acceptable to the moral communities to which the agent is responsible). There could then be a new exchange of reasons concerning whether and how the agent should be punished, but this is beyond the *moral responsibility exchange* itself, as McKenna rightly says.

The warrant for blame tracks this commitment to blame; it is the result of every excusing and exempting condition being exhaustively considered. I would go further and say that moral responsibility (for wrongdoing) itself tracks the commitment to blame. This is expressed in the following thesis:

MORAL RESPONSIBILITY AS APTNESS TO REACTIVE COMMITMENT: An agent S is morally responsible for act A with *agent meaning* Φ if she cannot in principle successfully defend the standpoint T (there being no winning strategy, and not just because she fails to find the winning strategy) that Φ-ing is morally permissible under conditions C and that C obtains in this case, in a critical discussion where the antagonist is some moral community M that S is responsible to and that has the authority to demand that S defend T.

If S fails to give a successful defence then there is a dialectical obligation for M to blame and this is enough to say that a reactive attitude can be attributed to the moral community and that the community's moral practices express that reactive attitude. This is not a conceptual analysis of responsibility, but more like a decision procedure for attributing responsibility; on analogy with a kind of linguistic or argumentative competence, we evaluate the moral competence and performance of the agent by evaluating a dialectical situation governed by the rules of a critical discussion.

From this point of view, then, blame does require the having of a reactive attitude and is the expression of that attitude after all, as long as it is remembered that

it is the community that blames and has the reactive attitude, and that a community can be said to have a reactive attitude even if nobody in the moral community actually has that reactive attitude. Furthermore, I have argued that there are reasons why even a member of the moral community who is quite aware of this obligation may nevertheless not carry out this obligation because believing, rightly or wrongly, that he does not have the authority or to be in the right relation to be the one to blame the agent. The community may then be morally blameworthy for failing to blame the agent, even in circumstances where no member of the community could have performed the speech act of blame (or at least not felicitously, because lacking the authority). When it is the injured party who blames then it is only a genuine case of blame if he blames on behalf of the community and has authority granted him by the community to blame. This does not prevent it from being the case that the blame expresses the resentment he is undoubtedly feeling and not only the attitude attributed to the community.

This is roughly in accordance with the interpretation of the reactive attitudes theory that claims that to be responsible is to be appropriately held responsible held by McKenna, Watson, and Wallace. It differs though in one possibly controversial respect: the conditions C are not universal but may vary within the functional parameters set by Strawson's discussion of excuses and exemptions.

What of the force of blame? Blame was presumed to have a backward-looking aspect. This, the reactive attitudes theory says, is because blame involves the reactive attitudes, and in contradistinction to objective attitudes towards wrong-doing that are only forward-looking, reactive attitudes are backward-looking, referring to a moral demand that has been violated. This seems fair enough. The concern, though, is that this is not really reflected in the notion of commitment, because one can have a commitment and be completely unaware of having the commitment; the commitment is not an attitude of any kind. Thus, to say that there is a commitment to blame, that blame is warranted, is not to say that any member of the moral community actually blames the wrong-doer. In one way this is a strength, for it means that an agent cannot escape being blameworthy simply because nobody knows what she has done or because of other facts about would-be blamers; it avoids the Problem of the Stranger, for example, and other cases where connecting responsibility too closely to relationships and to emotions seems to have the unfortunate result that lack of a relationship or emotion seems to let the wrong-doer off. In a certain sense, once the *moral contribution* has been made, which is when the action is performed, the rest should follow quite formally, and although there may be, in the *moral address* stage, an

actual person actually expressing feelings of resentment, and there may be, in the *moral account* stage, actual verbal exchanges of excuses or apologies, it does not seem that there needs to be — the warrant for blame does not depend on the protagonist actually taking the winning strategy but only that there be one, which is one reason why it need not be the agent who actually acts as the protagonist in the argumentation stage. However, the real issue is whether, when there is an attitude corresponding to a commitment, expressing that attitude has the communicative function that the *moral responsibility exchange* confers on it; it does, and this does not seem to be negatively affected by the fact that there could be a commitment without the attitude.

It has also been said that part of the force of blame is to make a demand. What is this demand? It is not to avoid violating the originative demand — it is too late for that. I have argued following McKenna that it is a demand for a moral account of the agent's violating the originative demand, and as such must refer to that demand and to that violation. In this way, it is both forward-looking and backward-looking.

It is only the moral community as a whole, I contend, that can actually make such a demand and that can actually blame. When a member of the moral community makes such a demand or blames, they do so implicitly on behalf of the moral community also; it is the agreed-to norms and expectations of the moral community that provides the warrant, but not necessarily the authority. The details of who in the community can make particular demands is complex: a wife may be entitled to make demands of and have expectations of her husband that another may not be entitled to, and if, believing themselves not to be so entitled, that person may retract the reactive attitude they have towards the husband, and it seems that it may be appropriate for them to do so. Is this a case where being responsible (as the husband certainly is) and appropriately being held responsible come apart?

Possibly so if holding responsible is to actually have an attitude or make a judgment, but not if holding responsible is a dialectical obligation. A reason to modify your own reactive attitudes is not always a reason for the moral community to modify the reactive attitude that can, by virtue of the agent's failure to give an acceptable moral account, be attributed to the moral community. This means that, although it is always as a member of a moral community that you blame and that you are provided with a warrant for blame, there are factors that affect your warrant that do not affect the warrant of the moral community as a whole. The moral community has a responsibility to blame that is distinct from your own.

This can be so even if you do, in fact, have the authority to blame. You may be such that when you blame, this counts as the moral community blaming, and yet reasons that would make you retract your blame would not make the community as such retract blame. You do not carry out your dialectical obligation to blame, and this itself is a demand that has been violated and that initiates a new moral responsibility exchange. So, for instance, you may not blame a particular wrong-doer because at the time when you could have done so you were caring for your sick child.[18] This may well be considered a good enough reason — it seems to depend on the moral community in question. One can certainly imagine Strawson's finding such a reason acceptable, but we must be careful of cultural bias here, and should not rule out moral communities that would not find this an acceptable excuse.

Let us suppose that you are not blameworthy or morally responsible for not blaming. However, I would say, following Copp (2007), that the moral community is morally responsible for not blaming nonetheless, because the excusing reason is not such as can excuse the community. Whether or not this is right, the important point is that there is a moral demand and dialectical obligation that does not disappear because of such reasons, for if there were then there would be nothing for either the community or the one in authority to be morally responsible for. Neither the lack of authority nor the hypocrisy of would-be blamers provides an escape route from being morally responsible.

REFERENCES

Barth, Else and Erik Krabbe. (1982). *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation*. Berlin: Walter de Gruyter.
Bell, Macalester. (2012). "The standing to blame: a critique". In D. Justin Coates and Neal A. Tognazzini (Eds.), *Blame: its nature and norms* (pp.263-281), New York: Oxford University Press
Copp, David. (2007). "The collective moral autonomy thesis". *The Journal of Philosophy* 38 (3), 369-388-29.
Eemeren, Frans H. van, Rob Grootendorst and Francisca Snoeck-Henkemans, et al. (1996). *Fundamentals of argumentation theory.* Mahwah, New Jersey: Lawrence Erlbaum Associates
List, Christian and Philip Pettit. (2006). "Group agency and supervenience". *Southern Journal of Philosophy* 44 (S1), 85-105
McKenna, Michael. (2012). *Conversation and responsibility*. New York: Oxford University Press
Macnamara, Coleen. (2012). "Taking demands out of blame". In D. Justin Coates and Neal A. Tognazzini (Eds.), *Blame: its nature and norms* (pp.141-161), New York: Oxford University Press

---

[18] Hypocritical blame could come in here as well. A blamer may, because he thinks he would have behaved in the same way or has behaved in the same way in the past, retract their reactive attitude. Here, though, I do not think that the reactive attitude would actually be inappropriate.

Pereboom, Derk. (2012). "Free will skepticism, blame, and obligation". In D. Justin Coates and Neal A. Tognazzini (Eds.), *Blame: its nature and norms* (pp.189-206), New York: Oxford University Press

Sher, George. (2012). "Wrongdoing and relationships: the problem of the stranger". In D. Justin Coates and Neal A. Tognazzini (Eds.), *Blame: its nature and norms* (pp.49-65), New York: Oxford University Press

Strawson, P. F. (2008). "Freedom and resentment". In *Freedom and resentment and other essays*. New York: Routledge, 1-28

Wallace, R. Jay. (2010). "Hypocrisy, moral address, and the equal standing of persons". *Philosophy & Public Affairs* 38 (4), 307-341

  - (1994). *Responsibility and the moral sentiments.* Cambridge, Mass. and London, England: Harvard University Press

Watson, Gary. (2004). "Responsibility and the limits of evil: variations on a Strawsonian theme". In Gary Watson, *Agency and answerability* (pp.219-259)*,* Cambridge: Cambridge University Press

DAVID BOTTING is Postdoctoral Research Fellow at the ArgLab of the research institute IFILNOVA attached to the Universidade Nova de Lisboa. His research interests are argumentation, logic, fallacies, and reasoning, on which he has had many publications in the journals Argumentation, Argumentation and Advocacy, Informal Logic, Cogency, and elsewhere.

R I A