

**David Botting**

Universidade Nova de Lisboa

## BEING REASONABLE

**Abstract.** Where do the rules of critical discussion get their normative force? What kinds of norms are involved? Unreasonable behaviour in the critical discussion – e.g., continuing to assert the contradictory of a proven standpoint, performing some action pragmatically inconsistent with a proven standpoint, or the same with regard to the starting-points agreed to in the opening stage – is liable to moral sanction. Thus, a moral/ethical norm is involved and the rules must have a moral force. Pragma-dialectics as it stands does not seem to account for this moral force. I will attempt to fill this gap in pragma-dialectical theory.

*Keywords:* pragma-dialectics; normativity; strategic maneuvering; collective intentionality; co-operation

### 1. Why be reasonable?

The pragma-dialectical concept of a critical discussion is defined as being aimed at a reasonable resolution of a disagreement on the merits of the better argument. To some extent, then, the normative force of the rules of critical discussion derives from the fact (assuming it to be a fact) that they are *problem-valid*, i.e., that they converge on better arguments when followed and diverge therefrom when violated.<sup>1</sup> Ultimately, the problem-validity of the rules must be verified empirically, but by showing that the fallacy types are all explicable in terms of violations of the rules, and conversely by using the rules to predict new fallacy types, the pragma-dialecticians can cogently argue that there is good reason for provisional acceptance of the rules. The normativity here explicated is procedural and adequately accounts for the correctness of the given argumentation.

However, we expect that the rules should have a moral normative force as well. This is a datum that this procedural notion does not account for. Perhaps by taking this as a starting-point it might be felt that I am taking

too much for granted. I should stress that I am not impugning the procedural notion described above, nor am I impugning that this notion supports the truth of the kind of ought-sentences like “You ought to A and you ought not to B” that, when analysed, turn out to be elliptical for the instrumental norm “If you want to achieve X, and A-ing is a means to X-ing, then you ought to A.” Nor am I concerned with what other norms may be involved. I claim only that there is a moral normative force and a moral “ought.” I do not intend to give a sophisticated argument but support it simply by linguistic intuitions about the kinds of things we say, and the sense in which we mean the evaluative terms that we use, when we criticize those whose conduct in the critical discussion we deem unreasonable.

If the normativity is procedural alone, what prevents a person from saying “Yes, I agree that the standpoint has been shown to follow from propositions I agreed to. But I still don’t believe it and will continue to act as if the critical discussion had never taken place. I have a hunch that there is a mistake somewhere, and perhaps I was simply wrong to agree initially to the starting-points I agreed to.” There is something natural about this; we often feel something to be true, whether through bias or through intuition, that we do not easily abandon even when compelled to admit that we have lost the argument. We are inclined to say something like “I feel that eventually time will show that I am right.” When we have lost the argument it does not seem right to say that we are morally compelled to believe what we nonetheless cannot bring ourselves to believe; strange to say, losing the argument only rarely brings us to change our mind, and this intransigence, in itself, is not necessarily unreasonable.

Nonetheless it does seem right to say that the standpoint, although not believed by us, is among our commitments, and that it is wrong to do or say anything that publicly commits us to anything incompatible with that commitment, and that when this wrong conduct is criticized by the words “You ought not to have done that!” this ought, and the wrongness it picks out, is moral. I think that the explanation for this is that the resolution resulting from correctly carried out argumentation is morally binding, and I will explain later why this is so. This also underlies the political principle that a person is obliged to abide by the results of a democratic process freely entered into and properly conducted<sup>2</sup> even when the person disagrees with the results. The loser’s attitude in this situation should be that of *acceptance*.

I am not saying that it is necessarily morally wrong all-things-considered to break the law or other agreement, since some laws are morally execrable

and in such dilemmas breaking it may be the right thing to do, and some agreements may conflict or be seen by the actor as conflicting with other moral principles of greater weight. What I am saying is that such a person is morally criticizable by the group whose attitudes the resolution reflects or whose joint activities may have to be modified or abandoned when the agreement is broken; the law-breaker is morally responsible to that group, and although he may successfully defend himself against the group such that it may not be moral for them to criticize him, there is a moral case for him to answer and a moral obligation to engage in argumentative discussion of its merits. One cannot rule out the possibility that the group will criticize him anyway, quite irrespective of his reasons or the merits of his defence, but such a sanction would not be a moral sanction. Thus, when I say that an actor is morally responsible, this is not the end of the story but the start of a new discussion. It is only this judgment of moral responsibility or accountability that I wish to establish in this paper.

But perhaps the would-be loser might prefer to avoid having to accept the standpoint by breaking the rules of the democratic process or critical discussion before the standpoint is established. What is to stop him from saying “Yes, I did agree to these things, but now I want to opt out. Of course, you may not wish to have further dialogue with me, but on what grounds can you morally sanction me for opting out?” In general it is not simply instrumentally counter-productive to break the rules or to *derail* the critical discussion by fallacious strategic maneuvering – it is morally wrong to do so. Again, this kind of unreasonableness seems to be a moral failing.

One might put the question “Why be reasonable?” However, Popper showed a long time ago that this question was self-defeating, for if one was not prepared to accept reasonableness in the first place then it is futile to give reasons for being reasonable; one cannot without circularity provide a rational justification of reasonableness. One does not reason with a creature driven purely by its appetites. If you want an animal to come indoors you do not give it an argument but tempt it inside with food. Similarly, it is pointless to give reasons to someone for whom reasons carry no weight, irrespective of how good those reasons are. Instead there must be, says Popper (1986, 225–31), a faith in reason he calls the *rational attitude*. It is a condition of critical discussions that the participants have a *rational attitude*; this is tied up with their *critical aim* for a reasonable resolution of the discussion on the merits of the better argument.

So, let us put the question a different way, as “Why *continue* to be reasonable?”

## 2. Why continue to be reasonable?

A speaker has voluntarily agreed to co-operate with another with the *critical aim* of bringing about a reasonable resolution. So, a preliminary answer might be that the *problem-validity* of the rules guarantees that they converge on a goal that all participants have voluntarily agreed to. Again, the procedural notion seems to provide part of the answer.

However, it does not provide all of the answer, for what is to stop a speaker from simply withdrawing from the critical discussion, especially if he sees that he is likely to lose the argument? Yet surely we would not consider this reasonable; we would criticize and in some cases even sanction it. We would say “You ought not to have done that” and mean by the use of this “ought” that the agent is morally accountable, that he at least owes us an explanation or justification of his actions. It seems that one should still be able to hold the speaker to the commitments he or she engendered during the critical discussion and sanction the speaker for performing public actions whose entailed commitments are incompatible with those earlier commitments.

What does pragma-dialectics have to say here? Probably that the critical discussion has become *derailed* because the speaker’s *rhetorical aim* of winning the argument has overtaken in importance the *critical aim*; but this describes the situation without accounting for it. Perhaps that higher-order conditions for critical discussion are not satisfied; but this judgment does not seem to have the normative force we are entitled to expect. It is morally criticizable to put winning the argument ahead of reaching a reasonable resolution once one has voluntarily entered into a critical discussion, and one cannot exempt oneself from such criticism simply by opting out.

I propose that the explanation for why we should continue to be reasonable is tied to the fact that critical discussion, and probably dialogue in general, is a co-operative activity even when agonistic. I would identify our intuitions about unreasonableness with intuitions about behaviour in such co-operative activities. Perhaps not all co-operative activities, but those in which behaviour counts as being *jointly intentional* are like this.

So, a first attempt at a more sophisticated answer is that the resolution aimed at by the critical discussion is not a case of each speaker individually accepting the standpoint but of the group consisting of both speakers *jointly* accepting the standpoint. According to Tuomela’s (2001) concept of *acceptance* a member of a group may be compelled by virtue of his membership in the group and perhaps by roles he may occupy within the group to *accept* propositions he may *qua* individual believe to be false.<sup>3</sup> When the

role occupied is defined according to formal rules such as in a hierarchical organization, and the derivations of what they accept is likewise defined – such as a politician whose role it is to represent the opinion of the majority of his constituents and derives her acceptance from the individual acceptances of the constituents and her accepting taking a simple majority as the proper procedure for deriving acceptances in this case – Tuomela calls it a *positional belief*; she *accepts* their opinion (although she may believe it to be mistaken) without necessarily believing it. To take another example, an inferior may *accept* the view of a superior whose authority over him is a part of what functionally constitutes the organization in question by deriving it from what he believes to be the individual belief of the superior and his accepting the chain of command as the proper procedure. What should be accepted depends ultimately on mental states of the group members, but not necessarily on their beliefs about the proposition in question, as we shall see. Accepting what should be accepted involves drawing this inference.

These derivations/inferences are what Pettit (2004) calls ‘premise-based’ procedures because the correct answer is derived from premises the group accepts by a procedure the group accepts rather than from the members’ prior beliefs about the proposition in question.<sup>4</sup> It is such premise-based procedures that seem to give the correct result in so-called ‘discursive dilemmas’ and ensure rationality at the group level. Consider a tenure committee. Individual  $M_1$  believes that candidate D excelled in research and teaching, but not in service. Since  $M_1$  believes that D did not excel in all three areas, for these beliefs to be deductively closed and hence rational,  $M_1$  must also believe, and does, that D should not be granted tenure. Individual  $M_2$  believes that D excelled in research and service but not in teaching, and so agrees with  $M_1$  that D should not be granted tenure. Similarly, individual  $M_3$  believes that D excelled in teaching and service but not research, making it unanimous that D should not be granted tenure. These beliefs are all rational on the individual level.

However, the majority believed  $p$  that D excelled in teaching (individuals  $M_1$  &  $M_3$ ), a majority believed  $q$  that D excelled in research (individuals  $M_1$  &  $M_2$ ), and a majority believed  $r$  that D excelled in service (individuals  $M_2$  &  $M_3$ ). They also all believe *not-g* that D should not be granted tenure. The problem is that they also believe the biconditional ( $g$  if and only if ( $p$  and  $q$  and  $r$ )), i.e., that tenure should be granted if these propositions are true; furthermore, the majority do believe of each conjunct that it is true, so if they believe the biconditional, then they should also believe D should be granted tenure in order for the group to be rational. Otherwise, the beliefs of the group will not be deductively closed. But we achieve deductive

closure of the group's beliefs if we choose a premise-based procedure of taking this biconditional as a procedure so that the group's belief about  $g$  is not based on their individual beliefs about  $g$  but are based instead on the members' sets of beliefs about the premises  $p$ ,  $q$ ,  $r$ , and the rule ( $g$  if and only if ( $p$  and  $q$  and  $r$ )). If they accept this last as a valid rule, then the procedure to ensure group rationality is to take some function of  $p$ ,  $q$ , and  $r$ , e.g., a majority vote, and to apply the rule (List and Pettit, 2006, 95–99). However, although it seems like the committee itself should accept  $g$  and, there being a *prima facie* obligation to accept personally what the group accepts jointly, every member of the committee should accept  $g$  and not act incompatibly with it, it does not seem necessary that any of them believe  $g$  or even should believe  $g$ .

The premise-based procedure is quite similar to the critical discussion in so far as the proposition accepted by the group can be shown to follow from propositions the members of the group jointly accept (i.e., the material starting-points agreed to in the opening stage, themselves formed from what the members of the group individually believe or accept) and an algorithm (i.e., argumentation schemes or inference rules agreed to in the opening stage) jointly accepted to generate the correct inferences. The role of the protagonist is then to demonstrate to the antagonist that the standpoint is necessarily derived from these starting-points when the algorithm is applied correctly. Showing that this – combined with the rules of critical discussion – ensures rationality at the group level provides additional force to the rules and grounds criticisms of irrationality, and suggest that it is irrationality at the group level that is morally criticizable, whereas irrationality (in the sense explained above) at the personal level may in fact be necessary for rationality at the group level.

What was it about choosing and following a premise-based procedure that made it jointly intentional? Tuomela says that a joint acceptance must satisfy an authority requirement – the right social and normative circumstances, as given by formal and informal rules, must be present. These rules define what the position-holder has the authority to do in virtue of the position he holds and are analogous to the speech-act conditions of performatives. If the position-holder exceeds their authority, then this does not count and any propositions he accepts in this mode are not binding on or, properly speaking, attributable to the group; the group has not acted intentionally.<sup>5</sup> For instance, if one of the members of the tenure committee was not properly elected to the committee in the first place then the committee does not grant tenure irrespective of what attitudes can be attributed to its members; what they may think are acceptances attributable to the group

may in fact be only their own individual acceptances, or to put it another way, they are collective but not joint. In speech act terms, the position-holder's acknowledgment of the proposition misfires in much the same way that a baptism misfires if not performed by an ordained minister. Anticipating slightly, I would say that no shared commitment is created by such an acknowledgment, although a personal commitment is; this does not rule out the possibility that all the participants in the critical discussion believe it to be shared and may legitimately take themselves to be entitled to moral sanctions when, as a matter of fact, they are not – this just repeats the point that what they may think are acceptances attributable to the group may in fact be only their own individual acceptances. In these cases, the sanctions can be analysed as speech acts that themselves misfire for much the same reason as the original acknowledgment. These are not *moral* sanctions, despite the fact that everybody involved thinks they are, including the one sanctioned.<sup>6</sup>

Underpinning these rules (or speech act conditions) are “general constitutive rules concerning the purposes and functions of the collective” and “proper social norms specifying his social roles” (Tuomela, 2001, 8). If the authority requirement is satisfied and the procedure followed then whatever issues from it is jointly intentional and has normative force. When a proposition is jointly accepted then it qualifies as a mutual belief and group-binding. Group-binding beliefs are: a) normative (Tuomela, 2001, 11) in such a way that a group-member who behaves in a way incompatible with this acceptance may rightly be morally sanctioned by the group;<sup>7</sup> and b) modelled in pragma-dialectics as *shared commitments*. Thus, what I have said so far can simply be seen as a gloss on the pragma-dialectical model rather than a criticism or modification of it.

Intersubjective agreement on its own (e.g., acceptance that is collective without being joint) does not do the job, because one can break such agreements without necessarily incurring criticism. The notion of intersubjectivity being used must be taken as possessing this kind of joint intentionality. The main task of this paper (tackled in section 3) turns out to be to show how exactly conditions sufficient for joint acceptance of all the dialogical commitments can be brought about in a broadly pragma-dialectical theory. Whereas in the cases considered by Tuomela the authority requirement is satisfied, when it is satisfied, prior to the act, here this authority has to be produced as part of the procedure itself.

As a preliminary the notion of *acceptance* has to be linked to the dialectical notion of *commitment*.<sup>8</sup> The speech act of assertion may be the externalization of an acceptance or a belief, or perhaps may even be a lie if

the context cannot distinguish between happy and unhappy acts of assertion<sup>9</sup> – the speech act whether happy or not generates a commitment to the proposition. However, this commitment is not shared and hence should not be added to the shared commitment-store unless the acceptance is jointly intentional. It might be thought that if we personally accept something that we do not believe then this can only be because we are members of a group that accepts it, whereby all acceptance is joint acceptance. This is probably true. It certainly seems true for *positional beliefs*. If she wishes to convince her constituents that a particular proposition follows from their views the politician may engage them in critical discussion and in the context of this critical discussion her personal acceptance is a joint acceptance, and by virtue of this the proposition qualifies as a mutual and group-binding belief attributable to the group. But what she personally accepts is not a joint acceptance unless accepted jointly with the particular antagonist in question, or perhaps it might be better to say that it does not produce a joint commitment; hence, if she tries to convince a different antagonist and engage him in a critical discussion her personal acceptance is not joint acceptance because not accepted jointly with him. It is not inconsistent for a person to have a commitment (or, for that matter, moral obligations) relative to one group or critical discussion that conflicts with a commitment relative to a different group or critical discussion. In each case she is morally responsible to the particular group with which she is currently in dialogue and can be held accountable by them. Of course this may create moral conflicts but this does not alter the fact that her current dialogue partners are entitled to morally criticize behaviour that is unreasonable from the point of view of the particular critical discussion, at least in so far as they are entitled to demand that she give an explanation or justification.

In the opening stage what adds a particular commitment to the shared commitment-store is a pair of dialectical moves, the first of which, performed by the protagonist, is the assertion of a proposition, and the response to which, performed by the antagonist, is acquiescence to the proposition. Here, as is generally the case, the protagonist is the *operative member* of the group, yet he does not act jointly with the antagonist unless the antagonist responds by showing at least passive acceptance, in the absence of which the proposition asserted does not amount to a joint commitment and is not added to the shared commitment-store. What determines whether a joint commitment has been produced is less a question of the kind of speech act performed or who is performing it, but of the kind of intentionality with which it is performed. The protagonist making an assertion does not yet produce a joint commitment and the acceptance is not yet attributable to

the group. The antagonist must take his turn, although he need not do anything when it is his turn. So although in one sense it is only the protagonist who acts, he does not without the co-operation of the antagonist accept jointly with the antagonist the proposition he asserts, or produce a joint commitment to this proposition. If the antagonist does not question the proposition when it is his turn he implicitly accepts the proposition and, provided that the protagonist and antagonist satisfy the authority requirement, makes the acceptance a joint acceptance and the proposition a joint commitment. This is what happens in the opening stage with the result that all the starting-points are joint commitments. If instead the antagonist questions the proposition then the proposition does not become a joint commitment until the protagonist has shown the proposition to follow from starting-points and inference schemes that are among the shared commitments; in other words, he shows that acquiescence to the proposition is rationally demanded for the group's acceptances to comply with the rationality constraint of deductive closure. In models that allow arguers to have commitments that are not shared, propositions that are asserted but not jointly accepted go into the speaker's own commitment-store.

What I hope to have shown is that the kind of things we say about group intentionality, especially the moral obligations and entitlements that we intuitively judge to be connected to group intentionality, are the same as we wish to say about critical discussions and the normative force of their rules, and conversely that critical discussions are already very similar to the kind of co-operative activities that are analysed under the rubric of group intentionality and shown by List and Pettit to satisfy group-level rationality constraints. Together these seem to show that breaking the rules is irrational at the group-level and that this irrationality is unreasonable, and that this unreasonableness is a moral failing, at least in so far as one can be morally obliged to account for it.

The observation that dialogue is a co-operative activity has been made before many times, for instance in Cohen and Levesque (1991, 956–57), and that dialogue games can be viewed, in the phrase of Hulstijn (2000, 3), as “recipes for joint action” whose constraints are basically those of a joint action is not new either. What these accounts have in common, however, is a continued adherence to a belief-desire-intention model.<sup>10</sup> We can integrate the belief-desire-intention model underlying the analysis of group intentionality with the dialectical model by noting that acceptance can be modelled as a commitment because externalized as an assertion, and joint acceptance by joint commitment when the assertion is acquiesced to, and joint commitments are already what the shared commitment-store contains. Group

intentionality in the dialectical form of joint commitment is the source of the normative force it is the object of this paper to explain. The main task now is to show exactly how this group intentionality emerges, show under what circumstances it is possible to opt out of the critical discussion, and what happens to our moral obligations when we do. I will also argue that the moral normativity is connected to commitments rather than beliefs, desires, or intentions. What distinguishes my efforts, then, from those others is that it attempts to deal entirely in terms of the commitment model.

This is something that, because of strategic maneuvering, even pragma-dialectics as it currently stands does not do. The pragma-dialectical concept of strategic maneuvering is not only inadequate but is actually inconsistent with the rest of pragma-dialectics. The basic idea behind strategic maneuvering is to attempt a rapprochement with insights gained from the study of rhetoric. The basic point is that it is not inconsistent with having the *critical aim* of reasonable resolution to have also the *rhetorical aim* of winning the argument, and once we accept this, it is also acceptable to prefer one of several equally correct arguments on the grounds that it is more likely to be persuasive or is in some sense better tactics. Unreasonableness only follows when the *rhetorical aim* conflicts with the *critical aim* or rather (since some participant's *rhetorical aim* ultimately must conflict with the *critical aim*) where the speaker wishes to drop the *critical aim* in preference to the *rhetorical aim*.

The problem is that whereas the *critical aim* is modelled by virtue of being a condition of the speech act complex of argumentation, the *rhetorical aim* is not modelled at all. There is no commitment that seems to be associated with the *rhetorical aim* or speech act of which it seems to be a condition or part of the content. What pragma-dialectics says about strategic maneuvering is actually framed in terms of a belief-desire-intention model and not a commitment-based model at all. Pragma-dialectics talks of commitments in order to avoid ascribing mental states to individual agents, and then it introduces strategic maneuvering by ascribing the desire to win. Note also that this ascription cannot be to a speaker defined merely as the occupier of a functional role as it should in a purely dialectical model but only to a genuine self-identical intentional subject whose mental states persist over time. In short, strategic maneuvering is at odds with a commitment-based model. Hence, there is already a theoretical need within pragma-dialectics, all moral considerations aside, to incorporate everything into a single model.

Above I identified the *rhetorical aim* with opting out of the critical discussion. It is plausible that one's desire to be reasonable may diminish if one sees that one is not likely to win the argument, and although it

is perhaps uncharitable to suppose that this is each speaker's sole reason for entering the critical discussion in the first place, the goal of reasonable resolution may cease to be of value to the speaker should the *rhetorical aim* no longer seem achievable. What we need in order to model *rhetorical aims* are speech acts aimed at forming and disbanding groups. The theory of such acts should show how the group intentionality with its corresponding moral normativity and the moral norms associated with opting out derives. Luckily, there already is such a theory. In a number of papers Cohen and Levesque have defended a view of group rationality explicated in terms of speech acts. They also view dialogue itself as such a co-operative activity. In the next section I will outline Cohen and Levesque's theory and show how I intend to use it.

### 3. Joint commitments and speech acts

The two main speech acts in Cohen and Levesque's theory are *request* and *confirm*. These form a pair. Bill *requests* the cooperation of Fred in building the wall. Fred *confirms* that he will help Bill. By virtue of a successful request-confirm pair, Bill and Fred now have a *joint persistent goal* to build the wall that persists until either the wall is built, one of them comes to believe that the goal of building the wall cannot be achieved, or the reason one of them had for having the goal of building a wall no longer applies. Because the goal is a psychological attitude defined in terms of a belief-desire-intention model there is no goal and hence no team if one participant's reason for having the goal dissipates, but the participant is nonetheless morally obliged to make this fact mutually known, and this goes analogously for the other two exit conditions; each member of the disbanded team is morally obliged to inform the other members if any of these situations occur.

It is questionable, however, whether this is adequate. Gilbert (1997, 67–68) reasons about an example of two people who decide to talk a walk for half an hour. After ten minutes one of them says “I do not intend to go on.” The other, Gilbert says, does not have to accept this: when one has jointly formed an intention with another to do some action, one cannot unilaterally choose to break off that action, and if one should do so then the other is entitled to rebuke you irrespective of whether you make your intention mutually known – making this known is not to deny that there is a joint intention but to act in defiance of it. This does not seem to be captured by Cohen and Levesque's account. I hold that it is essentially the same intuition that is operative here that is captured in our intuitions about unreasonableness.

Also, Gilbert goes on, it is not impossible that the other party might reply “I did not intend to go on for half an hour either!” Gilbert (1997, 67–68) concludes that here there is a shared intention despite neither party intending to complete the intended action. This goes against most analyses of jointly intentional action where for the action to be jointly intentional it must be performed in the correct intentional state; for instance, on Cohen and Levesque’s account when Bill requests that Fred performs some action with the goal of building the wall, it is necessary that Fred not only does the action but that he does it in the mental state of having what they call a *weak achievement goal*; briefly, that he does it *as a response to Bill’s request that he do it*. Furthermore, by relativizing this goal to a further condition (the third of the exit conditions described above) allowance is made for them to be able to cease to have the shared intention and cease to form a team, without, they point out, this meaning that they no longer have any moral obligations to each other. But these moral obligations do not go far enough; what they have allowed for is only an obligation to make the falsity of the relativizing condition mutually believed, and the inadequacy of this is clear from what Gilbert says above.

Nonetheless, I think that Gilbert has drawn the wrong conclusion. She starts from certain moral intuitions and then designs her account of shared intention around it. The conclusion that she should draw is that the moral intuitions do not stem directly from the joint intentionality but from commitments. In fact, she does use the phrase ‘joint commitment’ but it is not the dialectical notion of commitment that she has in mind. The same point seems to emerge from the following criticism made by Allwood, Traum, and Jokinen (2000, 880): if Fred only pretends to join the group then there is no team and no joint intention despite the fact that Fred voluntarily (if insincerely) issued the *confirm*. As they point out, Fred would have moral obligations by virtue of the external commitment he takes on when he performs this speech act irrespective of whether he was sincere when doing so.

Cohen and Levesque’s theory does not properly account for the ethics of the situation, but this does not mean that it is wrong as regards to its analysis of intentionality itself. I am inclined still to believe that acting intentionally whether by an individual or by a team is a matter of having the right psychological attitudes. But being a member of a team and thereby criticizable by other members of the team is not a matter of having the right psychological attitudes but simply of having performed the request-confirm pair.<sup>11</sup> As far as integrating the Cohen-Levesque theory with pragma-dialectics goes, then, the issue is resolved anyway by these ethical issues in favour of the commitment-based model of the latter.

There also seems to be a formal problem with Cohen and Levesque's account that makes it questionable as an analysis of intentionality itself: it is part of the content of the *weak achievement goal* that an actor must have in order for his action to be jointly intentional that this goal be relativized to the reasons *for which the one issuing the request* had the goal (Smith and Cohen, 1995, 5). But this seems like a mistake, since surely Fred can have his own reasons for having the goal, and it is when these reasons lapse and not when Bill's reasons lapse that Fred would desire to opt out and is obliged to make this mutually known. This condition, then, should not be part of what Bill attempts to get Fred to believe when he attempts to form a team with Fred – it does no harm for Fred to believe that Bill is 'fanatical' about his goal, for Bill will be obliged to make it known to Fred that Bill no longer has the goal if this comes about. In fact, there is a formal proof that, assuming sincerity, the requester must have the 'fanatical' goal they call the *persistent goal* (Smith and Cohen, 1995, 5–6). The relativizing condition, then, should not be part of what is mutually believed or part of what Bill is trying to get Fred to mutually believe when issuing the request.

On the current analysis the team exists by virtue of sharing a *joint persistent goal*, but this has to be changed to so that the speech acts themselves form a team by virtue of producing *joint commitments*. This amounts to dropping the sincerity assumption. One's attempts to form a team by a *request* then always succeed when followed by a *confirm* even if the one issuing the *confirm* was insincere or even misunderstood the *request*.

No relativizing conditions should be externalized or included among the shared commitments; therefore, the obligation is not to assert that a specific relativizing condition is false but that whatever is the relativizing condition is false. For example: Bill's relativizing condition for building the wall is  $q$ . If Bill ceases to accept  $q$  his obligation is not to assert "Not  $q$ " or "No commitment  $q$ " because there never has been a shared commitment to  $q$ . Nor can Bill simply assert "No commitment  $p$ " where  $p$  is an optative proposition "Let's build a wall together" or some such because this simply violates Bill's commitments. What it seems Bill must do is *request* a dialogue (probably a negotiation dialogue) whose standpoint is "No commitment  $p$ ." This implies that he places no value on building the wall. It is then up to Fred whether he *confirms* or *refuses* the *request*.

I have emphasized relativizing conditions because it is here that those *rhetorical aims* discussed earlier naturally seem to fit. Instead of building a wall, suppose that Bill wants Fred to co-operate with him in a critical discussion. The *joint persistent goal* of such a discussion is the *critical aim*,

namely a reasonable resolution of disagreement. It seems quite plausible that one's reason for having this goal is because one expects to win, and in the absence of this expectation perhaps one would cease to have the goal. On the original Cohen-Levesque theory this *rhetorical aim* would be modelled firstly by being part of the content of the speech act of requesting and secondly by being part of the *weak achievement goal* as a consequence of which the participant must be acting if his action is describable as intentionally co-operating and not just a happy accident. In short, the *rhetorical aim* is a relativizing condition. There should be no *joint commitment* to the *rhetorical aims* of the participants, and indeed it would make no sense if there were because both participants want to win and so the shared commitment-store would become inconsistent.

So what happens when the relativizing condition becomes false, e.g., Bill comes to believe or expect that he will lose the argument? According to Cohen and Levesque, there is an obligation to make this fact mutually known but the team is disbanded nonetheless. I agree that there is such an obligation but I agree with Gilbert against Cohen and Levesque about the team being disbanded unless this is agreed to by other members of the team. It should not be so easy to disband the team even by explicitly asserting that one's relativizing condition has become false, and making it known that you no longer value the goal does not amount to retracting the commitments that you have already engendered up until now. The commitments persist and every member of the team can be held to accept them.

What should happen if, for instance, Bill (the protagonist) wants to opt out of the critical discussion once it has begun, is that he has to issue a second request to disband the team. As mentioned, this might be modelled as putting "No commitment  $p$ " as the requested task and amounts to expressing indifference to reasonable resolution of the disagreement. In an unmixed discussion this is all that Fred (the antagonist) wanted anyway so he may well *confirm* this request, although he is by no means obliged to do so, for by doing so he frees Bill of his obligations so that Bill does not act unreasonably when he acts as if  $p$  were true, nor can Bill be held to account for doing so. In a mixed discussion, however, where Fred's *rhetorical aim* is commitment to *not- $p$* , he is likely to *refuse* the request. This seems to lead to an impasse, since Bill may well prefer to disagree than continue. Thus, failure of the relativizing condition will often lead to failure of the possibility condition, and the request to disband can be seen as the means of making this known. At this point the goal becomes impossible to achieve and the team is disbanded in this way, but the joint commitments persist and Bill can still be held to them though Fred, I will now argue, cannot.

Suppose now that Fred acts in a way incompatible with one of the joint commitments.

Is he morally criticizable? Not by Bill, who disqualified himself from criticizing Fred when he requested the disbanding of the team, because this amounted to an admission by Bill that he did not care sufficiently about being reasonable. This is a case where a *tu quoque* such as “How can you expect me to value continuing to be reasonable when you have made known that you yourself do not value it?” seems perfectly apt.

Remembering that only jointly accepted propositions get added to the shared commitment-store, that to be joint an authority requirement must be satisfied, and that it is because of the normativity of such group-binding beliefs (i.e., joint commitments) that acting incompatibly with the shared commitments is morally criticizable, it follows that to close a critical discussion the standpoint must be asserted or retracted by a member who has the authority to be the *operative* member of the group. A member who has requested a disbanding has ceded this authority and cannot be the one to make this dialectical move. So if the protagonist asserts the standpoint now it does not become a shared commitment even if it has been shown to follow from the starting-points after all. Being unable to close the critical discussion is equivalent to being unable to sanction Fred’s behaviour, for by Bill’s own actions the rules are prevented from converging on the better argument, or indeed on anything – the critical discussion has become fatally derailed. Fred’s response is that Bill does not have the authority, so the critical discussion remains open unless Fred chooses to close it.

On the other hand, if it is Bill who behaves incompatibly with a shared commitment then he is morally criticizable by Fred. However, Fred also gives up that right when he behaves incompatibly with a shared commitment. In such a situation neither party is morally criticizable by the other because neither is morally responsible to the other or to the now-disbanded team.

What is asserted but not jointly accepted should, I think, go into a personal commitment-store. This is required for three reasons. Firstly, the speech act itself, even if it cannot be attributed to the group as it is when jointly intentional, naturally seems to produce a commitment and it does not seem to make sense to simply ignore these commitments. Secondly, it seems morally criticizable to violate one’s own commitments as well as one’s joint commitments, even if this criticism can often be rebutted on the grounds that violating one’s own commitments has become necessary as a result of what one has been compelled to accept *qua* member of the group and these latter commitments have not been violated. I am not saying that it is morally wrong to violate one’s own commitments, or to violate the

groups' commitments, for that matter. What I have been arguing is that the group (in other words, one's dialogue partner) can hold one responsible for violating a group commitment in that it is entitled to demand an explanation or justification, and one is morally obliged to concede to this demand; what I would like to say now is that I think there must be some sense, and still a moral sense, in which one is open to criticism for violating one's personal commitments. Remembering that irrationality at the individual level may be necessary for rationality at the group-level, the justification that one might be obliged to give may appeal to the fact that violation of the personal commitment was required in order to avoid violation of the groups' commitments, and thus that the group is not entitled to any sanction. However, I am inclined to think that there is still an obligation to give this justification if it is demanded. Thirdly, although the rules of pragma-dialectics as they currently stand do not allow appeal to propositions that are not intersubjectively agreed starting-points an arguer should, I think, be able to say that he has such and such reasons for believing  $p$  even when these reasons are not starting-points, and these reasons should go into a personal commitment-store that can however be inspected by another; such an arguer has more credibility than someone who gives no reasons at all, and is at least epistemologically (though perhaps not morally) criticizable if he continues to believe  $p$  having ceased to believe in these reasons. This is what Sinnott-Armstrong (1999) calls *arguer justification*. By giving reasons that can be attacked, by making himself vulnerable to criticism, the arguer expresses the *rational attitude*.

#### 4. Conclusion

The positivists argued a long time ago that 'ought' sentences can often be rephrased as statements of instrumental rationality: given goal  $G$  and a means  $M$  of achieving it, we 'ought' to  $M$ . The normativity that pragma-dialectics accounts for is for this kind of 'ought' – given the goal of a reasonable resolution and a means of achieving it (which, by definition, is to follow rules that are *problem-valid*) we ought to adopt that means. Thus, when the *problem-validity* of the rules they put forward is established (which, as I have said, is an empirical question) it is established simultaneously that one ought to follow them in the described sense of 'ought.'

There is more than one kind of 'ought' however, and the one described does not say why, for instance, one should have the goal  $G$ , or why one should not abandon that goal. A partial answer is that the goal has been agreed to by all concerned parties. But here agreement to and continued

adherence to a common goal is assumed from the outset. Granted that we have set out with the desire to be reasonable, why must this continue? Why does prior agreement to this goal make it the case that one 'ought' not to break the agreement? Of course there may be conflicts of moral duties where one 'ought' to break the agreement in order to fulfil a more pressing obligation. Nevertheless, the 'ought' is there and remains there, and we have to account for this.

The intersubjective agreements that take place in the opening and closing stages of the critical discussion must have a normative force that is more than instrumental and that pragma-dialectics does not currently account for. It must be moral. The rules of pragma-dialectics express this moral force: they are not the source of it, nor is joint commitment the source of the moral force of the rules (I am not claiming that the rules are normative because jointly accepted), but it does explain why they have moral force. When one violates one of these rules it is not just bad sportsmanship but a moral failing because sabotaging a jointly intentional activity. This is reflected by the nature of the commitments: the intersubjective agreement is not to be given a simplistic distributive analysis as "the protagonist accepted *p*" and "the antagonist accepted *p*" but as "the protagonist and antagonist jointly accepted *p*." This does not require the impossibility of *some* distributive analysis; the premise-based procedures of Pettit amount to just such an analysis. The externalization of joint acceptance in speech acts by one with the authority to do so leads to a joint commitment, and the joint commitment is binding on the protagonist and antagonist, at least relative to the critical discussion they are engaged in. This 'jointness' explains the moral force of the rules and answers the question "Why continue to be rational?"

Arguers entering into a critical discussion need not be supposed to both belong to some community in which they hold particular social roles with clearly specified duties. On the contrary, they must deliberately form a community between themselves when initiating a critical discussion; thereafter the dialectical model itself divides duties clearly between the protagonist and the antagonist. The right social and normative circumstances mentioned in the authority requirement do not precede the critical discussion but are coeval with it, and are brought about, I have argued, by a request-confirm pair as discussed by Cohen and Levesque. This also, as Cohen and Levesque show (despite shortcomings in their analysis), would lead inevitably to the protagonist and antagonist having certain moral obligations to each other.

I have argued that it should not be as easy to opt out of the critical discussion as is implied by Cohen and Levesque. The person who wants to

opt out is obliged to make this known by issuing a second request, but in doing so he cedes the authority to close the critical discussion, and with it the authority to act as a moral judge. Even if, at the end of it all, the protagonist has shown the standpoint to follow from the starting-points, still the critical discussion is not closed, because the standpoint is not in this case a joint commitment, which is a way of saying that when the protagonist acts Tuomela's authority requirement is not satisfied. Whoever has not tried to opt out is the operative member and only for them is the authority requirement satisfied, and this only for so long as they do not themselves break any of the rules.

Even if all this is unsatisfactory, I hope to have left the reader with one insight into the reasonable resolution of a critical discussion that I think outweighs all the inferences I have drawn from it, which is that when there is such a resolution, this is not just a matter of a protagonist and an antagonist having a commitment to the same proposition, but of a *shared commitment* to the standpoint; that is to say, the group consisting of both the protagonist and an antagonist *jointly accept* the standpoint. This may be different from what they as individuals believe or accept; for instance, the member of the tenure committee may accept, in order to have deductively closed acceptances at the individual level, that the applicant does not deserve tenure, while accepting, in order to have deductively closed acceptances at the group level, that the applicant does deserve tenure. This does not involve the member in any kind of epistemic vice, as it would if we attributed to the member belief that the applicant does and does not deserve tenure, or oblige him to believe something he takes to be false. Furthermore, even if he is right and the standpoint is false, he is still morally accountable to the group, but this does not mean that he is morally wrong. The issue is over moral responsibility and accountability, and not of moral right or wrong. The man who follows what is socially acceptable is often not morally superior to the man who follows his conscience, but he is more open to moral criticism and less able to defend himself against it.\*

## NOTES

<sup>1</sup> Note that this only needs to hold in general and does not need the better argument to win in every single case. That the rules may sometimes have a non-normative result when followed (that is to say that bad arguments sometimes win) does not by itself show the rules themselves to be non-normative; to show this would require showing that iteratively applying the rules would not eventually or usually give the normatively correct result (that is to say that good arguments win in the long run).

<sup>2</sup> Philosophers and deliberative democrats like Habermas and Rawls have taken great pains to spell out exactly what it is for such a process to be 'properly conducted'.

<sup>3</sup> In contrast with belief, acceptance is an action and performative rather than experiential and dispositional. Belief when attributed to a group means only that the members *accept* it as the view of the group and as potentially a reason for action, and not that all the members believe it is true, or even that any member believes it is true (as some examples of discursive dilemmas show).

<sup>4</sup> The law tends to use premise-based procedures. For instance, a jury is not really asked whether or not they believe a defendant to be guilty but whether certain conditions legally sufficient for a guilty verdict have been met. If any of the conditions are not met in the view of some member of the jury then that person has to return a verdict of “not guilty” (which is not, of course, to say that they always do so). This procedure is stipulated and usually passed on to the jury in the judge’s closing instructions. A certain proportion of the jury have to return “Guilty” in order for the jury as a group to return a guilty verdict. This procedure also is stipulated.

<sup>5</sup> For instance, if a minor functionary of an organization signs a contract on ‘behalf’ of the organization, then such a contract is not legally binding if, in so signing, the functionary broke the rules delimiting his authority.

Tuomela also requires satisfaction of an intentionality requirement. Basically, he is saying that there are act-types that are intrinsically joint, and that joint acceptance is one of them. I am not convinced by this requirement but it is not the place to argue the issue here and it would not affect my argument either way.

<sup>6</sup> Perhaps we should not always say that these are not moral sanctions, because it seems like a person is entitled to criticize another if that person contradicts one of their personal commitments and not only a shared commitment. The moral force involved seems to be different, however, and weaker.

<sup>7</sup> Gilbert (1997, 67) expands on the kind of normativity involved:

- a. Each participant has an obligation not to act contrary to the shared intention.
- b. Corresponding to these obligations are rights or entitlements of the other parties to the appropriate performances.
- c. If one participant does something contrary to the shared intention, the others have a special entitlement to rebuke that person.

She calls this the *obligation criterion*. There is a broad consensus that there are moral norms of this kind, differing only in details, such that I do not intend to defend these claims or argue for them further, nor does it matter for my purposes which account of group intentionality is correct as long as it accounts for these moral norms (Bratman’s is the only account of joint intentionality that I know of that denies any moral normativity). On the contrary, my intention is to take them as the explananda and account for them.

<sup>8</sup> Their close connection is mentioned also in Walton & Johnson (2011).

<sup>9</sup> In some contexts indicators such as a sarcastic tone of voice or evidence that the speaker stands to gain personally from being taken to be asserting something makes the insincerity too obvious to be charitably interpreted as making an assertion at all. Also, if a speaker is self-consciously lying with a genuine intent to deceive then it is questionable whether he is arguing at all since he is no longer attempting to reach a reasonable resolution on the merits of the better argument, but here I will assume that he is arguing and hence that a commitment is produced by virtue of the assertion. Certainly, such a speaker is still under the kind of moral obligations at issue – it can be no excuse when he acts contrary to a starting-point that he was lying when he agreed to that starting-point.

<sup>10</sup> For example, Hulstijn (2000, 7) says: “The smallest recipes for joint action are precisely the exchanges described by dialogue game rules. On the other hand, plans and goals may function as a semantics for dialogue game rules. They motivate the illocutionary point of initiating a game and explain various aspects of cooperativity in dialogue.” A semantics of this kind is foreign to dialectical models and is in fact irrelevant for determining how

well a given dialogue fits the model, that is to say, for determining the correctness of the dialogue.

<sup>11</sup> The irrelevance of psychological attitudes extends only to the judgment of moral responsibility, i.e., that there is a case to answer, and is not meant to imply that in arguing the merits and demerits of this case the psychological attitudes of the actors involved cannot relevantly be appealed to. Obviously, the motives of actors are highly relevant to deciding any moral issue.

\* The author would like to acknowledge financial support from the FCT Portugal through grants SFRH/BPD/77687/2011 and PTDC/FIL-FIL/110117/2009.

## R E F E R E N C E S

- Allwood, J., Traum, D., & Jokinen, K. (2000). Cooperation, dialogue and ethics. *International Journal of Human-Computer Studies*, 53, 871–914.
- Cohen, P., & Levesque, H. (1991). Confirmation and joint action. In *Proceedings, International Joint Conference on Artificial Intelligence* (pp. 951–957) (IJ-CAI '91). Morgan Kaufman Publishers, Inc. Available at <http://ijcai.org/Past%20Proceedings/IJCAI-91-VOL2/PDF/051A.pdf>
- Gilbert, M. (1997). What is it for *us* to intend? In G. Holmström-Hintikka & R. Tuomela (Eds.), *Contemporary Action Theory, vol 2: Social Action* (pp. 65–85). Dordrecht/Boston/London: Kluwer Academic Publishers.
- Hulstijn, J. (2000). Dialogue games are recipes for joint action. In R. Cohen (Ed.) *Proceedings of the Fourth Workshop on the Semantics and Pragmatics of Dialogue* (Gotalog'00). Available at [http://www.researchgate.net/profile/Joris-Hulstijn/publication/2331454\\_Dialogue-Games\\_are\\_Recipes\\_for\\_Joint-Action/links/02bfe5112130a2e5d8000000.pdf](http://www.researchgate.net/profile/Joris-Hulstijn/publication/2331454_Dialogue-Games_are_Recipes_for_Joint-Action/links/02bfe5112130a2e5d8000000.pdf)
- List, C., & Pettit, P. (2006). Group agency and supervenience. *Southern Journal of Philosophy*, 44 (S1), 85–105.
- Pettit, P. (2004). Groups with minds of their own. In F. Schmitt (Ed.), *Socializing Metaphysics* (pp. 167–93). New York: Rowman and Littlefield.
- Popper, K. (1986). *The open society and its enemies: Vol 2*. London: Routledge.
- Sinnott-Armstrong, W. (1999). Begging the question. *Australasian Journal of Philosophy*, 77 (2), 174–91.
- Smith, I., & Cohen, P. (1995). Toward a semantics for a speech act based agent communication language. In *Proceedings of CIKM Workshop '95 on Intelligent Information Agents*. Available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.8767&rep=rep1&type=pdf>.
- Tuomela, R. (2001). Collective intentionality and social agents. In *AI Conference IMF*, Toulouse, France, 2001. Available at <http://www.helsinki.fi/teoretiskfilosofi/personal/Sintonen/Explanatory%20Essays//tuomela.pdf>
- Walton, D. & Johnson, R. (2011). Introduction: Special issue on Charles Hamblin. *Informal Logic*, 31 (4), i–iv.