



# Moral criticism, hypocrisy, and pragmatics

Y. Sandy Berkovski<sup>1</sup> 

Accepted: 14 July 2022 / Published online: 16 August 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

A good chunk of the recent discussion of hypocrisy concerned the hypocritical “moral address” where, in the simplest case, a person criticises another for  $\phi$ -ing having engaged in  $\phi$ -ing himself, and where the critic’s reasons are overtly moral. The debate has conceptual and normative sides to it. We ask both what hypocrisy is, and why it is wrong. In this paper I focus on the conceptual explication of hypocrisy by examining the pragmatic features of the situation where accusations of hypocrisy are made. After rejecting several extant views, I defend the idea that moral criticisms are best understood as moves in an agonistic or hostile conversation, and that charges of hypocrisy are attempts to prevent the hypocrite from gaining an upper hand in a situation of conflict. I finish by linking this idea to frame-theoretic analysis and evolutionary psychology.

**Keywords** Hypocrisy · Blame · Pragmatics

## 1 Introduction

A good chunk of the recent discussion of hypocrisy concerned the hypocritical moral address where, in the simplest case, a person criticises another for  $\phi$ -ing having engaged in  $\phi$ -ing himself, and where the critic’s reasons are overtly moral. The debate has conceptual and normative sides to it. We ask both what hypocrisy is, and why it is wrong. In this paper I focus on the pragmatic features of the situation where accusations of hypocrisy are made. I will argue that moral criticisms are normally made in the situations of rivalry over certain assets. In such situations the rivals’ moral wrongdoings decrease their chances to obtain those assets. Accusations of hypocrisy are aimed to prevent this eventuality by pointing out similar wrongdoings, committed either by the critic himself or, generally, by a person affiliated with the critic. Towards the end I will also say a few words about the hypocrite’s moral fault of deception.

The plan is this. In Sect. 2 I reject a popular idea that hypocrites implicate or presuppose certain falsehoods about themselves. In Sect. 3 I reject another popular

---

✉ Y. Sandy Berkovski  
sandy.berkovski@gmail.com

<sup>1</sup> Department of Philosophy, Bilkent University, 06800 Ankara, Turkey

idea that hypocrites' utterances are infelicitous for the lack of their moral standing. In Sect. 4 I describe Daniela Dover's challenge to the anti-hypocrisy norm. I respond to it in Sect. 5 where I outline my own view of moral criticisms and hypocrisy. I further defend it in Sect. 6.

## 2 Hypocrisy: between implicature and presupposition

In its simplest form a hypocritical address has the form:

- (1) i. At  $t$ :  $A$ :  $B$  was morally wrong to have  $\phi$ -ed.
- ii. (At  $t' \leq t$ :  $A$  is  $\phi$ -ing.)

Under these conditions,  $A$  is supposed to be rightly accused of hypocrisy. Of course the surface form of  $A$ 's utterance may vary: "Morally speaking,  $B$  was wrong to have  $\phi$ -ed", " $B$  shouldn't have  $\phi$ -ed", and more specific expressions like " $B$ 's  $\phi$ -ing was an ugly betrayal". Common to all these forms is a criticism directed at a particular person and made in overtly moral terms.

Therefore, the exchange in full would be this:

- (2) a.  $A$ :  $B$  was morally wrong to have  $\phi$ -ed. (Moral address)
- b.  $B$ : But you  $\phi$ -ed yourself! (Riposte)

Just like the address, the riposte too is highly schematised. In actual conversations it comes in all sorts of forms. I only assume that we can identify in it an appeal to the past  $\phi$ -ing by  $A$ .

Right away, a few disclaimers. I will ignore the more complex cases where you may be accused of hypocrisy about  $\phi$ -ing even if you haven't  $\phi$ -ed in the past.<sup>1</sup> For the most part I will also assume that the conversation is between  $A$  and  $B$ , and that a riposte to  $A$ 's criticism is made by  $B$ . In some discussions this assumption is intrinsic to the argument, but I'm making it for simplicity only (I return to this issue in Sect. 4). As well, hypocrisy is sometimes ascribed to the person's character, thoughts, and non-verbal behaviour. You may, e.g., be judged a "hypocrite" if you regularly go to church whilst violating every other article of the Christian creed in private. As I argue in Berkovski (2022), there are meaningful connections between these ascriptions of hypocrisy and the hypocrisy of moral address. Here I focus on the latter and ignore the former.

Now, what is  $A$ 's fault in (2)? There *must* be a fault, since our ripostes are rich with outrage and indignation and often couched in moral terms.<sup>2</sup>  $A$  is a hypocrite, and hypocrisy is morally repugnant. This anti-hypocrisy norm is shared by most writers on the subject.<sup>3</sup> Later in Sect. 4 we'll look at an attempt to challenge this norm. For the time being let's grant its validity.

<sup>1</sup> Subjunctive hypocrisy: Piovarchy (2020:10–11).

<sup>2</sup> This brings me to another disclaimer, that criticisms not couched in moral terms are outside the scope of our discussion. This includes, e.g., the emotionally charged, angry complaints of the kind sketched in Wolf (2011).

<sup>3</sup> See, e.g., Wallace (2010), Rossi (2021), and every writer in the debate cited here, with the possible exception of Dover (2019a) discussed below. But whatever the consensus in the current debate over blame

If *A* is wrong morally, this is because of what he said: if *A* stayed silent, he wouldn't have stood accused of hypocrisy.<sup>4</sup> So to understand *A*'s moral fault we must first understand what hypocrisy *is*, and this means to understand better the features of his utterance. So there are two related questions, the pragmatic and the moral. Something in the hypocrite's utterance is faulty, and the riposte is meant to identify, somehow, that very fault. The fault ultimately turns out to be moral, but the hypocrite manifests it by pragmatic means—i.e. by making the utterance of criticism.

Two further preliminary comments. (i) Precisely because we believe that *B* is trying to identify *A*'s fault we judge his utterance felicitous. Else we would have judged *B* uncooperative: without any warning he states a certain fact of *A*'s biography and changes the subject of the conversation. But in reality, if *A* replies, "How is this related? I'm talking about *you*", *B* would remark (correctly, we believe) that his riposte *was* in fact related to the original address, and that far from merely discussing *A*'s biography, he is trying to identify a fault with the initial utterance. (ii) In his riposte *B* doesn't simply point out that *A* is a hypocrite by definition. This would leave open the question why whatever *A* said—i.e. his "hypocrisy"—is morally faulty in the first place. It's the other way round: *B* identifies some problem with *A*'s utterance and on *this* basis concludes that *B* is at fault. Whether this counts as "hypocrisy" or not isn't the issue the riposte is meant to settle. Rather, it is we qua philosophical observers who describe *A* as a "hypocrite".

Our task, then, is to examine *A*'s original address with the view of explaining *B*'s riposte. Many theorists are attracted to the idea that a moral criticism is associated with the speaker's special moral commitment—namely, the commitment to examine one's own life with regard to  $\phi$ -ing. Thus, e.g., Jay Wallace argues that by uttering his criticism the hypocrite *A* incurs a "latent" commitment to self-scrutiny (2010:328). On the other hand, *A* hasn't honoured that commitment himself. This is a moral failure of partiality: *A* treats others differently from himself, imposes on them demands he does not impose on himself.

But how exactly is this self-scrutiny commitment associated with the hypocrite's utterance? Let's look at three possibilities. First, the commitment may be presupposed:

**Commitment presupposition** When *A* says "*B* was morally wrong to  $\phi$ ", it triggers the presupposition that *A* is committed to self-scrutiny with regard to  $\phi$ -ing.

Here "triggering" is understood in a semi-technical sense. The very use of certain expressions in an utterance is a reason enough for us to infer the presupposed truth-

---

and hypocrisy, might we not say generally that hypocrisy is often a salutary phenomenon, so far as it serves  
Footnote 3 continued

a useful social purpose? This, for example, is a view espoused in Mandeville (1924). Yet even Mandeville assumes at the outset that hypocrisy (or at least some of its forms) is in itself odious, although we may have to *tolerate* it for the sake of public utility. See, e.g., Douglass (2022:469).

<sup>4</sup> Well, can't there be hypocritical blame that has never been expressed verbally? Certainly. But I think it should be understood as a defect of character. It is a principally different form of hypocrisy to be given a different treatment. For example, a silent hypocrite wouldn't be guilty of deceiving others. He should instead be criticised for his own self-deception. See Berkovski (2022) for a discussion of public and silent hypocrisies and their links to self-deception. In any event, as mentioned earlier, this form of hypocrisy is outside the purview of this paper. The present task is to reconstruct the dynamics of a verbal interaction where hypocrisy is a driving force.

conditional content. If the content is in fact false, the utterance is infelicitous. Thus the very use of “morally wrong” or related expressions is a good reason for us to infer *A*’s commitment to self-scrutiny. The riposte would be analogous to the annoyed “hey wait a minute!” reactions to presupposition failures: the riposter is aghast at the fact that the speaker presupposes some content—namely, his commitment—that is uncontroversially false.<sup>5</sup> So the commitment presupposition passes the test of “uncontroversiality”: if *P* presupposes *Q*, then the assertion that *P* is felicitous only if *Q* is uncontroversial for the audience.<sup>6</sup> Therefore, if *Q* is controversial for the audience, then their reaction to *P* would be to challenge *Q* first. What’s more, they might use some device to indicate surprise or annoyance. The riposte “But you  $\phi$ -ed yourself!” is a way of expressing that surprise: the presupposition of the speaker’s commitment has been “generated” by the address, but it is controversial and is, therefore, challenged first. The moral address itself is infelicitous.

This argument is unconvincing. The annoyed reaction of the auditor may be attributed to a variety of reasons.<sup>7</sup> Here, we may interpret the auditor’s reaction in (2) as *moral* indignation. Someone who doesn’t care about hypocrisy or  $\phi$ -ing generally is unlikely to complain about the critic’s utterance. By contrast, when we deal with a real presupposition, then the complaint is sensible whatever the auditor may think about moral issues:

- (3) a. *B* was morally wrong to fornicate.  
b. What do you mean? *B* is a virgin!

Secondly, the commitment presupposition fails the survival under entailment cancellation: if *P* presupposes *Q*, then both *P* and  $\sim P$  should imply *Q*, where “imply” is taken in some informal sense.<sup>8</sup> Then consider:

- (4) a. (*A* says:) *B* was morally wrong to fornicate.  
b. (*A* says:) *B* wasn’t morally wrong to fornicate.  
c. *A* is committed to self-scrutiny.

Assume that the criticism does presuppose the commitment. But, if the speaker refuses to criticise *B* for  $\phi$ -ing, then we *can’t* assume that he is committed to self-scrutiny with regard to  $\phi$ -ing. A natural reason of his refusal, by Wallace’s own lights, is exactly the failure of the commitment. Hence the riposte would be out of place:

- (5) a. (*A* says:) *B* wasn’t morally wrong to fornicate.  
b. ?? But you fornicated yourself!

I conclude that Wallace’s version of commitment is not presupposed in the moral address.

Other theorists seem to think that the commitment is communicated. Lippert-Rasmussen (2020:667) attributes to Wallace the view that a hypocrite “implicitly

<sup>5</sup> Presupposition triggers: Levinson (1983:178–179), Tonhauser et al. (2013:75). “Hey wait a minute!”: von Fintel (2004:316).

<sup>6</sup> See Potts (2015:174), Simons (2006:359).

<sup>7</sup> See Tonhauser et al. (2013:81).

<sup>8</sup> See Simons (2006:358).

denies” the equality of persons. According to Fritz and Miller (2018:125), hypocrites “implicitly reject” the equality of persons. And Rossi (2018:562) simply says that hypocrites “communicate” a commitment to the norm of non- $\phi$ -ing. These formulations may suggest that hypocrites convey some false content to their audiences:

Commitment communication When *A* says “*B* was morally wrong to  $\phi$ ”, he communicates his commitment *M* to the norm such as “No-one should  $\phi$ ”, where the commitment to the norm *X* is (i) the judgement that the behaviour in accordance with *X* is good and (ii) the desire to promote *X* and to act in accordance with *X*. (Rossi 2021: 63ff)

On this view the riposte, again, is aimed at pointing out a deception: whereas the speaker has just suggested that he has a commitment *M*, the riposter points out, possibly by an implicature, that the speaker has no such commitment.

I think there is little evidence that *M* is communicated by the moral critic. Of course, it is not communicated as part of the truth-conditional semantic content: we don’t grasp the critic’s moral commitment merely by understanding the linguistic meaning of his utterance. Mightn’t it be communicated pragmatically as a conversational implicature? For this to happen, *M* must be something that the audience attributes to the speaker (or “calculates”) in order for the utterance to be intelligible conversationally.<sup>9</sup> We assume the speaker to be cooperative and follow Gricean maxims. Yet often, if all that the speaker has communicated was the semantic content of his utterance, then we would judge him at least not fully cooperative. Hence, in order to preserve the assumption that the speaker *is* in fact cooperative, we impute to him some further communicated content—namely, a conversational implicature. The utterance, we reason, was a way for the speaker to inform us about that content. He relied on our ability to conduct the inference and grasp the implicated content.

Now I take it that in general, *M* plays no such role in the conversation. Suppose *A* says, “*B* was morally wrong to flirt.” We are not required to impute to *A* a commitment to chastity (say) in order to make the utterance intelligible. The criticism is a legitimate conversational move regardless of the commitment *M*. If it turns out that *A* has no such commitment, and if he also assures us that he never meant to communicate it to us, we wouldn’t judge him an incompetent interlocutor. By the same token, a riposte is not a complaint about a conversational mishap caused by the hypocrite. Compare two exchanges:

*A*: I’ve run out of petrol.

*B*: There is a garage around the corner.

In this typical case *B* conversationally implicates that the driver can get petrol in the garage. If it turns out that the garage is closed, *B* can’t pretend that he *merely* meant to inform *A* that there is a garage there. To say so would show poor understanding of *A*’s previous remark (itself containing an implicature) or be an attempt at a poor joke. On the other hand, consider:

<sup>9</sup> Conversational implicature: Potts (2015:179). Difference between observing and flouting the maxims: Levinson (1983:104–105). Wallace himself floats the idea that the critic conversationally implicates that he hasn’t  $\phi$ -ed (2010:316, 334).

*A*: *B* was morally wrong to flirt.

*B*: But you don't care about chastity and such matters!

Here *A* may well insist that, whatever his own attitudes, *B* is still at fault, and that his criticism stands. He may be resisted on moral grounds, of course. But his criticism won't be judged pragmatically incoherent.<sup>10</sup>

There is, however, another interesting possibility. Even if *A* doesn't pragmatically "inform" us of his commitment, aren't we still entitled to infer it on the basis of his utterance? The utterance may be a "cue" like accent: to speak with a German accent is not to convey that you're a native German speaker, but it is nevertheless a cue of that. By the same token, though *A* isn't conveying his moral commitment, his criticism is a cue of it.<sup>11</sup>

Let's put this idea as follows:

**Commitment cue** When *A* says "*B* was morally wrong to  $\phi$ ", the audience is entitled to infer that *A* is committed to *M*. The inference does not rely on any pragmatic mechanism. In particular, the inference is warranted whether or not *A* wished the audience to recognise his commitment *M*.

There are two difficulties with this proposal. First, why would you think of a moral criticism as a reliable cue of *M*? Likely because you appeal to an inference to the best explanation: the commitment *M* would explain best the utterance of criticism. However, if hypocrisy is frequent, then a no less plausible alternative explanation would be that the speaker is a hypocrite lacking *M*. Or in Bayesian terms, given the familiar abundance of hypocrites, the criticism wouldn't raise the posterior probability of the hypothesis that the speaker has *M*. Under these conditions of abundant hypocrisy, the criticism ceases to be a reliable cue of *M*. So why to react with a riposte? If the criticism were reliable, then the riposte would have been a complaint against the speaker sending a deceptive cue. But if the cue is not misleading anyone, and if the audience is not entitled to infer *M*, the riposte's rationale is no longer clear. Yet in reality, I think, we are justified to issue a riposte also when we are certain of being surrounded by hypocrites.

Secondly, supposing now that the criticism is a reliable cue, the riposte still does not make much sense. If, e.g., an Australian utters, "It is raining outside" with a German accent, it is strange to protest, "But you aren't German!" It is a fact of daily life that people often manipulate their behavioural attributes to send deceptive cues. But even when you suspect deception, you don't normally engage with them by pointing it out.

There are exceptions. Sometimes the cue is so ostentatious that the whole point of the utterance (or of an action generally) seems to demonstrate the possession of the desired quality. Communication becomes pragmatic: e.g., the German ancestry of the Australian speaker may, in this case, be conversationally implicated. Similarly, the

<sup>10</sup> Since there is no presumption that *M* is conveyed pragmatically at all, it would be misleading to apply the usual tests of conversational implicature, like cancellability or reinforceability. These tests are designed to distinguish implicature from other kinds of non-at-issue content. But if *M* is not such a content in the first place, it may pass these tests trivially.

<sup>11</sup> A version of this view is in Shoemaker and Vargas (2021). I discuss it in more detail elsewhere. The terminology of "cue" is from Maynard Smith and Harper (2003) where it is contrasted with "signals". This contrast is loosely analogous to the one between meaning<sub>N</sub> and meaning<sub>NN</sub> of Grice (1957).

context may be set up in such a way that a moral criticism is best understood as *just* a means to convey the speaker's commitment *M*.<sup>12</sup> But as I said in the discussion of implicature earlier, I believe this is not the case generally: moral criticism does not function as a pragmatic vehicle of *M*.

Another exception is when only the people with the desired attributes are *authorised* to engage in a particular behaviour. It is urgent to point out the deceptive cue, since the agent has no right to behave the way he does. Thus, if the possession of *M* is a necessary credential to make the criticism, then the hypocrite is an impostor impersonating a "qualified" moral critic. His criticism itself is "invalid", and the riposte is meant to articulate this fact. I address this possibility in Sect. 3.

### 3 The pragmatic standing of authority

Many theorists have been attracted to the idea that the hypocrite lacks a certain "standing".<sup>13</sup> Schematically, according to these authors, *x* must meet a certain moral requirement *R* in order to have a "right" or "standing" to criticise (or "blame"), i.e. to utter some sentence *C*. But what does this right to utter *C* consist in, exactly? What happens, for example, when *x* hasn't met *R* and still utters *C*?

Whereas a lot has been said about the requirement *R*, *very* little has been said about how this right impacts the pragmatic situation of moral criticisms. And sometimes, when the authors did try to clarify the nature of that right, confusion reigned. Thus Riedener (2019:200–201) says, on one hand, that even if *A* has failed the requirement *R* and has no relevant standing, he *can* "really blame" *B*: "unauthorised blame is still blame". What *A* can't do, Riedener goes on to say in the same paragraph, is "express an essentially moral disapproval". Why? Simply because, having failed *R*, "he wouldn't be blaming"! So to blame "with authority" is to express moral disapproval. This requires *R*. But "unauthorised" blame is also blame.

This confusion may be remedied along the lines of Scanlon (2008:128). We should distinguish between an "impersonal" judgement of blameworthiness and a "personal" judgement of blame. The former is a vanilla moral judgement. The latter is a special judgement requiring *R*. Scanlon said nothing about the pragmatics of that personal judgement. Others went further and defended the following idea:

**Illocutionism** The speaker should possess the necessary qualifications, or "authority", or "standing", in order to perform an illocutionary act of criticism. Anyone who lacks the authority would fail to *make* the criticism.

This is the view in Cohen (2013a,c), Isserow and Klein (2017:198), possibly in Riedener (2019:196). A key piece of evidence for it comes from an utterance like (6):

- (6) You were wrong to flirt with her, but I'm not in a position to criticise you.  
(Cohen 2013a: 120–121)

<sup>12</sup> Special conditions for the cue to become pragmatic: Grice (1989:100–104).

<sup>13</sup> Some typical statements: Fritz and Miller (2018:125), Todd (2019:347, 357), Wallace (2010:332), Lippert-Rasmussen (2020:672), Rossi (2021:78), and further references below.

This utterance is puzzling: the speaker criticises in the first conjunct, yet immediately denies that he can do so in the second. Still, this is what we often say, apparently quite felicitously. The illocutionist has a ready explanation. (6) is felicitous, so the thought goes, because the speaker is entitled to assert something about flirting, as he does in the first conjunct. No special authority is required for that. But he is not entitled to criticise a particular person precisely for the lack of authority or “standing”. The ability to make sense of (6) is a major motivation of illocutionism. I’ll refer to (6) and the problem it generates as “Cohen’s puzzle”.

An initial worry about Cohen’s puzzle is whether the evidence is decisive, and our intuitions robust. The speaker in (6) may be willing to paraphrase his utterance as:

- (7) a. Someone might argue that you were wrong to flirt; myself, I’m on the fence.  
 b. Morally speaking/according to moral rules, you were wrong to flirt; myself, I don’t care about morals one way or another.

What seemed like the speaker’s own assertion is in fact a report of a statement, whether someone else’s assertion or a general moral claim. Since in these cases the speaker doesn’t even assert the wrongness of the act, the puzzle is dissolved. I suspect that this indeed is what sometimes goes on in the conversations of this sort, but arguably not always.

Supposing, therefore, that the puzzle is real, we are looking at this idea: a person may be able to assert that  $x$ ’s  $\phi$ -ing is wrong, but unable to perform *another* illocutionary act—namely, to “criticise”, or “blame”,  $x$  for  $\phi$ -ing. Hence the speaker in (6) asserts the wrongness of flirting, but does not, even by his own lights, criticise the interlocutor for that. What the illocutionary force of that act of criticism is remains *very* unclear. Isserow and Klein speak alternately of “condemning”, “sanctioning”, and “bestowing disesteem” (2017:198, 195, 193). Riedener, as already mentioned, settles for the “expression of the attitude of disapproval” (2019:195).

Beyond this basic unclarity, a general complaint against illocutionism is this. To entertain seriously the possibility that speech-theoretic concepts are adequate to resolve the puzzle, we must be able to classify criticisms as a particular kind of a speech act. Standard classifications contain two types of speech acts that require speaker’s authority for their performance. One is declarations. To declare war, to open a meeting, to marry people, one has to be invested with a socially recognised authority. Such authority enables the speaker to create new (social) facts. It seems trivial that the moral critic is not declaring or stipulating a new moral fact of the wrongness of  $\phi$ -ing. Among other problems, this view would rule out moral realism where moral propositions have truth values independent of people’s attitudes or beliefs.<sup>14</sup>

But, you might retort, no special authority is required for the assertion of the wrongness of  $\phi$ -ing. Since we now identify a further illocutionary act of criticism in the critic’s utterance, authority may well be required for that act. On Scanlon’s view, for example, to be able to blame you, as opposed to merely judging you blameworthy, I must be in a moral relationship with you unimpaired by my past wrongdoing like  $\phi$ -ing or worse (2008:175). So to blame Scanlon’s way I should be able to declare

<sup>14</sup> See, e.g., Shafer-Landau (2003:15) for a careful formulation of moral realism.

that our relationship has been impaired. A hypocrite guilty of past  $\phi$ -ing is not in a position to do so.

This retort is hardly convincing. For, by Scanlon's lights, when I *am* in the position to blame, I am not *stipulating* that you have impaired our relationship. Unlike the case of proper declarations, I'm not fashioning any new fact, like that the meeting is open, or that people are married. I'm rather registering an antecedent fact. Indeed, seen in speech-theoretic terms, Scanlon's blame is expressed in another assertion. A hypocrite's putative assertion of blame would simply be false, even if his assertion of blameworthiness be true.<sup>15</sup>

You might think that Isserow and Klein's idiom of "condemnation" is a better candidate for declarations. Their rich theory of moral authority too is well adapted to support this idea. But we can't push this too far. A moral expert gravely announcing, "I'm thereby condemning you for  $\phi$ -ing" does not *turn* you into someone officially condemned, so to speak. We don't think that your *status* has changed with that very announcement. Plainly the missing element is the institutional, conventional authority that moral experts can't claim.

The other type of speech acts tied to authority is directives like orders and commands. There is no consensus on this thorny subject. According to Bach and Harnish (1979:47) and Alston (2000:98), the speaker doesn't have to actually have the relevant authority for a command or an order to come off. To perform the directive successfully, the speaker need only to believe that he has the authority. Adopting this condition would defeat the whole purpose of our analysis: a hypocrite, at least a wholehearted one, would be able to criticise at will. His criticisms would "come off". Well, we may clear this hurdle by simply following Searle and Vanderveken (1985:91, 101) where felicitous directives do require a special authority of the speaker.

If charitably read, some of Isserow and Klein's remarks may align with this idea. We are told, e.g., that criticisms are "action-guiding": moral experts must be able to direct their audiences by making rules, adjudicating disputes, sanctioning violations, and reward obedience (2017:195, 197). Just like the necessary authority is a preparatory condition of a successful command, a moral criticism is only real when the speaker has a moral authority to influence the behaviour of the addressee.

As the details are scarce, it's hard to say conclusively whether this is Isserow and Klein's intended interpretation. It also doesn't sit well with their (and Riedener's) other remarks mentioned earlier. In any event, there is a good reason to reject it. A very basic problem is that utterances of criticism don't have the directive force. (8a) can't be paraphrased into (8a):

- (8) a. You were wrong to flirt.  
b. Do not flirt!

A descriptive content may be associated with the former, but not with the latter.<sup>16</sup> A standard move in the speech-act approach in such cases is to interpret a problematic utterance like (8a) as an indirect speech act that, in this instance, has the directive force despite being in the declarative mood. But of course, it is perfectly felicitous to

<sup>15</sup> This connects to the issue of the emotional component of Scanlon's blame. See Wolf (2011).

<sup>16</sup> Compare a related discussion in Portner (2016:606).

greet (8a) with the plain “No, I wasn’t!” This is an obviously infelicitous response to (8a). So we don’t understand the criticism as an indirect command.

Now, if a moral criticism is neither a directive, nor a declaration, can’t it belong in a class of illocutionary acts so far undiscovered? The chances of demonstrating this are slim, I think. When we say that for a given illocutionary act the speaker’s authority is, in some sense, “essential”, we mean that without that authority the act is “void”, doesn’t “come off”, doesn’t “happen”.<sup>17</sup> If you order me to shoot and I refuse, and if it later transpires that you had zero authority to order me so, I will claim that I didn’t *disobey* you: there was nothing there to disobey in the first place. If I send a paper to a journal, and an impostor pretending to be the editor declares that it is accepted, in reality the acceptance never happened, end of story.

Suppose now that a hypocrite *A* criticises *B*. Suppose that his hypocrisy, and with it the lack of the required credentials, is found out. One difficulty, mentioned lately, is to understand what it *is* that never happened; but put it aside. Usually, if *x* has no required authority, some *y* who has it would be called on to repair the act and to do what *x* has failed to accomplish—to open the meeting, to accept the paper, to order to shoot. So we might expect something to the same effect to be in our case. If *A* is found to be a hypocrite, then perhaps *C* should criticise *B* instead. But this, I think, is not how we look at the situation. Consider:

- (9) a. *A*: *B* was wrong to flirt!  
 b. *B*: But you flirted/fornicated yourself! Your criticism is null and void.  
 c. *C*: Well, I’ve always been chaste. And I’m telling you, you were wrong to flirt!  
 d. *A*: Very well, your criticism is real and legitimate.

This exchange is farcical at several levels. When put explicitly, the talk of the criticism being “null and void” is bizarre. Even if we speak occasionally of someone not having the “standing” to say this or that, we don’t take this idiom so seriously, as to believe that what the speaker says is literally null and void. If in the second half of (9b) *B* says anything meaningful at all, he says it elliptically. Perhaps all he says is, “I don’t care what you say anyway.” It is equally bizarre that in (9d) *B* should certify or accept that the criticism “happened”, at last. Just like the explicit talk of a criticism being void, the talk of a criticism being “certified” or “real” must strike us as odd.

Secondly, if *B* believes that *A*’s criticism was not real, then he must ignore the *criticism* entirely, as something that never happened. But he shouldn’t ignore *A*’s *assertion*: by assumption, the assertion was real. Plausibly, though, he doesn’t give *A*’s assertion much weight, precisely because *A* is a hypocrite. When it comes to *C*, the criticism is real. So how is *B* supposed to react, other than acknowledging the reality of that criticism? Well, conceivably he will pay more attention to *C*’s utterance, rather than *B*’s, and reconsider his behaviour accordingly. But then, it seems that the different reactions, as we understand them, to unreal criticisms as opposed to real criticisms are very much like the reactions to dubious assertions made by dubious speakers as opposed to the reactions to trustworthy assertions made by trustworthy

<sup>17</sup> The precedent here is, of course, the original discussion in Austin (1975:16).

speakers.<sup>18</sup> The authority of the speaker is required, not to *make* the criticism, but to make it more *effective*. We don't need a separate category of criticisms: we can make sense of the hearer's reactions using only the category of assertion that may be more or less believable or trustworthy. To put it differently: the speaker's authority, and the possibility of hypocrisy undermining it, come into play at the perlocutionary stage. They are not a factor at the illocutionary stage.<sup>19</sup>

I will return to this "perlocutionist" view in a moment. As for illocutionism, I conclude that the prospects of substantiating it are dim. This doesn't mean, I hasten to add, that we have resolved Cohen's puzzle. We must look for its solution elsewhere.

#### 4 Perlocutionism, hostility, and reform

To review the argument so far: we tried to understand the nature of the riposte. On the face of it, the riposte is an infelicitous change of subject. If in fact it *is* felicitous, that's because the riposter identifies a certain deficiency in the criticism and attempts to undermine it. Different authors offered different diagnoses of the deficiency (false implicature, illegitimate authority etc.), but we have argued that none of them is convincing. Along the way we have looked at Cohen's puzzle and argued that the speech-theoretic approach fails to resolve it. Hence any pragmatic account of hypocrisy must have at least two desiderata: to explain the felicity of the riposte and to resolve Cohen's puzzle.

Now it may be thought that there is a problem of a greater magnitude. Our discussion rested on an overarching assumption, shared by virtually all other writers on this topic, that moral criticisms are offered in a hostile manner. That is why the riposte purporting to undermine the criticism is felicitous in the first place. The charge of hypocrisy is a defensive weapon in the hands of the person criticised: one is always looking for a reason to disarm a moral criticism directed at oneself. The hypocrite is disarmed when he is told that his utterance is in some way deficient as a legitimate moral criticism.

Yet, how credible is the hostility assumption? Consider, for example, the perlocutionist view in Dworkin (2000). As Dworkin sees it, the moral critic's purpose is to "recruit agents into a common moral community" (187). It is to change the person's moral traits and dispositions. The right moral standing makes it more likely for the speaker to effect this change. The reason is that a moral criticism also expresses the speaker's disapproval of the addressee's action. And this very disapproval is the motive for change (187). We care about the opinion of others—but only of valuable others. Because the hypocrite is not someone whose opinions we care about, his criticism offers no motive for change.

But now, note some telling oddities in this explanation. First, by Dworkin's own lights, it is at odds with the general character of the exchange where a criticism is likely offered. Criticisms are assumed to be offered predominantly in a hostile

<sup>18</sup> Throughout I assume a traditional view of assertion, rather than a revisionist one like in Williamson (2000:249ff). Note too that the revisionist is likely to dispute the felicity of Cohen's puzzle (6) on a combination of moral and epistemic grounds.

<sup>19</sup> Curiously, Isserow and Klein (2017:199n16) mention the possibility of exactly this kind of explanation, but declare themselves unperturbed—wrongly, in my view.

manner: the critic gives the interlocutor “hard time”, as Dworkin (2000:187) puts it. Ripostes are similarly hostile. But if the criticism’s end is moral conversion through a disapproval expression, why can’t your disapproval be calm and friendly? Think of a lapsed Christian who comes to his priest to confess adultery. The priest wishing to “recruit” him back to the church would do well to adopt a non-confrontational, pedagogical approach. He would perhaps explain why chastity is in the parishioner’s own interest. It’s unclear, therefore, why giving the person hard time should be an intrinsic part of moral conversion.

Similarly, Dworkin’s account is at best incomplete in its explanation of the riposte. If the riposte is meant to inform the putative hypocrite that his judgement doesn’t count, what is the added value of hostility? Suppose that you criticise me for  $\phi$ -ing, and that you are so rotten that I don’t care a fig about your views. The riposte is meant to inform you about my attitudes. To this extent it makes sense. But should I be specially annoyed at your remark? You expressed your disapproval, gave me “hard time”. I explain to you why I’m indifferent to your disapproval. The usual emotional component of the riposte seems unmotivated.

Finally, consider third-person criticisms. Suppose I say, “Bill Clinton was wrong to fornicate.” Clinton is not in the audience that only consists of famously chaste people. I’m the only fornicator around. Evidently my utterance isn’t an attempt to change anyone’s behaviour. Generally too, in making a third-person criticism I can only rarely be understood as trying to change the audience’s behaviour. Even if the auditors are notorious fornicators, criticising someone else for fornication is too indirect to be of any use for changing the auditors’ ways.

Now Dworkin acknowledges that third-person criticisms needn’t aim at that: solidarity may be the goal, as I am trying to ally myself with chaste people (186). This seems *ad hoc*: why should third-person criticisms and hypocrisies be of a different kind than second-person ones? Why should a special moral authority be required to make a successful face-to-face criticism, but not necessary for a successful third-person criticism?

The upshot of these reflections is the idea that hostility and giving hard time are not easily aligned with moral reform and change. Hence a dilemma:

Dworkin’s dilemma Either: think of the moral criticism as an attempt to morally reform the hearer, discard the hostility assumption, and reassess the relationship between the critic and the target. Or: stick to the hostility assumption and do not think of criticisms as attempts to effect a moral change.

Daniela Dover has recently argued that we should embrace the first horn of the dilemma. She urges us to locate moral criticism within a “substantive” dialogue where criticisms are not elements of a “predictable stock ritual of condemnation and apology”, but rather occasions for reflection and self-improvement (2019b:401, 405). Dover credits Erasmus and the lesser known Andreas Karlstadt with engaging in this kind of “interactive criticism”. It imposes reciprocal requirements on the critic and the addressee. On one hand, the putative hypocrite is not only allowed, but encouraged to voice his moral disapproval of the addressee’s actions. This should be done in the ways that can contribute to a rational discussion where the criticism is articulated, not merely vented. The critic must conduct an “egalitarian” dialogue with the “fellow

inquirer”, the addressee. The addressee is required to do exactly the same, rather than brusquely dismiss the criticism on the grounds of hypocrisy. Only then “selves and social worlds are remade” (Dover 2019a:43, 47).

So the riposte, after all, would be a change of subject. Even if certain moral critics happen to be hypocrites, and even if the addressees are aware of their hypocrisy, the criticisms are still legitimate and should be taken at face value. We expect the conversation participants to interpret constructively the purpose of the criticism. Notice here that Dover’s argument is deliberately aimed at the conversations where the criticised person is directly addressed by the critic. Only then the addressee should be required to “listen” to what the critic has to say. Unlike Dworkin’s perlocutionism, however, this limitation isn’t necessarily *ad hoc*. It is part of the proposal that moral change and reform are possible in the course of an open, face-to-face dialogue where second-person criticism would be often be essential to the task. Third-person criticisms can’t play a comparable role.

I can’t do justice here to the richness of Dover’s ethical arguments. Instead, let me only restate her pragmatic argument as follows. In terms of the schema (1), we tell *B* criticised for  $\phi$ -ing:

- (10) Focus on your own  $\phi$ -ing. Even if you learn that *A*  $\phi$ -ed himself, ignore that, at least for now.

And to the original critic *A* we say:

- (11) Make sure to criticise *B* constructively. Don’t play the blame game. Suggest ways to improve *B*’s actions and character. Be ready to do your own soul-searching and improve yourself.

If the interlocutors stick to these maxims, they would be able to concentrate on their moral improvement. Hypocrisy charges would then be out of place, and ripostes infelicitous. The exchange would be respectful and mutually constructive, rather than hostile and destructive. This “constructive model” of moral criticism serves to justify the following pair of claims:

#### *Descriptive thesis*

As a matter of fact, there are exchanges spawned by moral criticisms that conform to the constructive model.

#### *Normative thesis*

Participants in moral exchanges should at least try to act according to the maxims of the constructive model.

Significantly, Dover never tells us how exactly moral criticisms are formulated. In the flagship examples of Erasmus’ and Karlstadt’s reactions to Luther it is not at all clear that their subject is a *moral* failure of either party. “Criticisms” there are in fact lengthy series of statements (and perhaps some non-linguistic acts) where the moral idiom doesn’t even appear. To consider the theses in such generality is far beyond the scope of this discussion (and I suspect that they would be unwieldy for a pragmatic analysis). Therefore, I will restrict both theses to more specific cases analogous to the schema (1) where criticisms are made in one or two utterances, and where the moral

vocabulary is explicitly used. Note too that this restriction may well go against much of Dover's motivation, so far as the substantive dialogue and lengthy interactions are contrasted with one-off criticisms.<sup>20</sup> Thus what we say about the constructive model might not fully apply to Dover's own understanding of it.

With these qualifications in place, a third desideratum for an account of hypocrisy is to defend the hostility assumption and to explain why Dover's constructive model is inadequate for characterising criticisms and ripostes. In the next section I outline such an account.

## 5 An agonistic model of hypocrisy

In this section I argue that both criticisms and ripostes are "agonistic" acts that occur in situations characterised by conflict over certain assets. It is this conflict that explains our intuitions about hypocrisy and our widespread aversion to it. After contrasting the agonistic model with Dover's constructive model, I offer a solution of Cohen's puzzle.

Consider this exchange:

- (12) a. *A*: Clinton was morally wrong to commit adultery.  
 b. *B*: But Gingrich fornicated too!

Its likely context is overtly political: behaviour and character of different politicians are debated and evaluated. It is understood that not all politicians are adulterers, and also that adultery generally lowers one's popularity. The criticism is a conversational move designed to group Clinton together with "bad" politicians, and then at some distant point to entice the audience to prefer "good" politicians over "bad" ones. The speaker is not a remote disinterested spectator. He has an interest in seeing Clinton fail, whether because he is a rival politician himself, or because he is affiliated with a rival political entity. Conceivably, (12a) is addressed to a political gathering or a TV audience, that is, the population that has a role in deciding Clinton's political fortunes. The same reasoning holds for a second-person address where we may imagine the speaker addressing Clinton at a public gathering.

These observations generalise. According to the present proposal, we locate moral criticism in an "agonistic" conversation where the participants are vying for an asset, whether a material resource, prestige, or love. It is mutually known that moral evaluations regularly affect one's chances in securing the asset. If a person is accused of a morally bad  $\phi$ -ing, then another person who hasn't  $\phi$ -ed is simultaneously elevated—merely because the former person's chances of securing the asset have decreased. The criticism, in other words, is a rhetorical instrument to recruit allies to your side in competition and conflict.

We now get a better view of a riposte like (12b). It is not meant to disqualify the putative hypocrite from making a certain kind of moral utterances or to dismiss his moral criticism out of hand. The riposter should certainly be content with either of these effects, but it is unclear, as I argued, how he can achieve them, or whether he can achieve them at all. Instead, we should think of his utterance as an attempt to prevent

<sup>20</sup> See Dover (2019a:38), Dover (2019b:400).

the context update proposed by the hypocrite by proposing *another* update—namely, the riposte proposition itself. The hypocrite proposes to change the common ground of the conversation—the set of propositions taken for granted by its participants—by adding to it a novel proposition, the content of his criticism. The riposter aims to add a further proposition. As a result, the criticism and the riposte balance each other out. If, in a simple case, we assume that *A* and *B* are vying for the same asset, and if their respective moral qualities are a factor in their rivalry, then the effect of the riposte is to deny the hypocrite *A* the advantage he wished to claim for himself and possibly damage his prospects vis-à-vis *B*. If the riposte is successful, the moral evaluations just advanced with regard to  $\phi$ -ing would be discounted in the present conversation. However, compared to the initial stage, the hypocrite's chances will be damaged. He will now be seen as someone guilty of a hypocritical deception. As far as the rivals' chances for obtaining the asset are concerned, following the riposte the hypocrite will lose some ground.<sup>21</sup>

How, then, does the hypocrite deceive? Not by communicating any false proposition or misleading the audience about his commitments. He deceives by proposing to update the common ground in a one-sided way. He deliberately ignores the facts of his own biography, generally the morally significant facts that are contrary to his interests in the given conversation. Thus he is able to manipulate the flow of conversation so as to make the faults of others mutually known. He designates them *as* faults and highlights them—whilst passing over his own faults. Thus we recover a very traditional understanding of the hypocrite's deception.<sup>22</sup>

There is an altogether more insidious and lasting aspect of this deception. The hypocrite is sometimes described as not taking morality “seriously”.<sup>23</sup> In a sense, though, he does take it seriously enough. He is adept at using the moral idiom to advance his interests. He is acutely aware of its vital role in situations of conflict. But so far as disinterestedness and impartiality are characteristic moral concerns, the hypocrite's behaviour is intrinsically a sham. Taken at face value, his moral criticism is a move in a debate governed by moral concerns. As things stand for him, however, it is a move in an agonistic exchange with a very different set of concerns. So to take the hypocrite's utterance at face value and to engage with him in a moral debate (say, by examining the moral qualities of  $\phi$ -ing) is to become an accomplice in his scheme. It is to lose sight of the actual goals of the exchange and to debate morality in a distinctively non-moral setup.

An upshot of these speculations is the claim:

<sup>21</sup> Could a *known* hypocrite plausibly issue a moral criticism? This interesting possibility will be addressed shortly. Secondly, what of a criticism made in private where *A* criticises *B* in a one-on-one interaction? Again, so far as *A*'s judgement is valuable to *B*, the latter has an interest in protecting *A*'s positive judgement of himself. More generally, we may appeal to the notion of “face” and say that also in dyadic interactions people protect their face and project a positive image of themselves (Brown and Levinson 1987). Moral criticisms damage the criticised person's face.

<sup>22</sup> See Shklar (1979:6), Szabados (1979:205), also Wallace (Wallace 2010:338). Whether the hypocrite *deliberately* deceives is an open question. See Dover (2019b:408–412) for a compelling survey of hypocrite's intentions.

<sup>23</sup> See Crisp and Cowton (1994:347).

Agonistic moralism Utterances of moral criticism and ripostes are appropriate, indeed expected in agonistic conversations.

Given Agonistic moralism, here is how to respond to Dover's constructive model. Participants in agonistic interactions cannot and should not be swayed by the maxims (10) and (11). The injunction against hypocrisy is an *intrinsic* feature of these interactions. Moral criticisms and ripostes to them accusing the critics of hypocrisy are both necessary for the participants conducting those conversations to achieve their goals. How, for example, in the schema (1) does *B* interpret the context in which the moral criticism was originally made—that is, how does he interpret *A*'s goals and his own? If the criticism was a means, say, to undermine *B*'s reputation, then to ignore *A*'s fault is to misunderstand the nature of the conversation.

Thus suppose that Newt Gingrich criticises Bill Clinton for adultery. If Clinton reacts by an earnest soul-searching and refuses to point out Gingrich's own adultery, then he fails to engage properly with the utterance. He *misunderstands* it. Gingrich's utterance was a move in a political conversation masquerading as a moral claim, and Clinton, by undertaking a scrutiny of his own behaviour in response, fails to participate competently in that conversation. Dover's normative thesis, in other words, seems to recommend to the addressee an irrational response. As for the descriptive thesis, clearly no support for it can be drawn from the fact of agonistic conversations.

There is a response to offer on Dover's behalf. We have imagined that moral criticisms occur in the middle of a conversation where the critic's agonistic goals are *already* manifest to the participants. What if these goals aren't initially known? The Gingrich-Clinton scenario is an outlier. Very few conversations, the response goes, are overtly agonistic. Often the critic's motives are sufficiently uncertain. And sometimes criticisms are made in overtly friendly conversations. This is just a gloss on the Descriptive thesis. So at least in those conversations the participants should follow the maxims of the constructive model. That's a gloss on the Normative thesis: a proper response to the criticism under those conditions should be an attempt at a moral improvement, not an accusation of hypocrisy.

To address this response I put forward the following claim:

#### *Agonistic transformation*

Utterances of moral criticism are a *pro tanto* evidence that the speaker's goals are in fact agonistic. They are sufficient to transform non-agonistic conversations into agonistic ones.

Suppose you turn on the TV where a panel host you haven't seen before in your life says a few unremarkable sentences and then issues a moral criticism of Clinton's adultery (to repeat: the criticism should be made in clear moral terms, perhaps using an explicit idiom of "moral wrongness"). Before the utterance you have no idea whether the speaker is a Republican party operative or a rival politician. For all you know, he may be a moral philosopher deeply concerned about morality. Then, I submit, agonistic assumptions are *created* from scratch. The speaker's purpose, hitherto opaque, is interpreted in this key: he is trying to gain some kind of advantage at Clinton's expense. The observer (who may or may not be in the audience) must revise his assumptions in such a way as to attribute to the speaker the hostile intention of damaging the criticised person's reputation.

Consider now a more difficult case still. Suppose that, at a certain stage of the conversation, you positively believe that your goals and interests are aligned with the goals and interests of your interlocutor. Let's say that you have a heart-to-heart chat with an old friend where you share with each other the problems of your life. And then he makes a second-person moral criticism, "You were morally wrong to  $\phi$ ." Right away, as far as you are concerned, the overtly friendly exchange is transformed into an agonistic one. Likelier still, you now believe that, from your interlocutor's perspective, the conversation has *always* been agonistic, and that he engaged in a "fabrication": he pretended to care about your interests, but in reality attempted to advance his own. Either way, you now hold that the interlocutor has certain interests incompatible with your interests, and his utterance is called to facilitate the satisfaction of his interests at the expense of your own (e.g., he aims to establish his moral superiority).

As a brief illustration, consider this exchange between two friends:

(Augustus Staveley:) "[I]f you marry her you will commit a great *sin*."

(Felix Graham:) "How *moral* you have grown!"

(Augustus:) "No, I'm not. I'm *not a bit* moral. But I know very well when a man is in love with a girl, and I know very well that you're not in love with Mary Snow. And I tell you what, *my friend*, if you do marry her you are *done for life*. There will absolutely be an *end of you*." (Trollope 1985:331, italics added)

Augustus' initial utterance is cast in moral terms which prompts an instant rebuke from Felix. A close friend, Augustus is now perceived as somewhat hostile. He immediately seeks to dispel any such impression by dropping the moral idiom and assuring Felix that he has his best interests at heart.

A few comments are necessary. First, in defending the claim of agonistic transformation I do not suggest that *every* conversation where a moral criticism occurs must already be or become agonistic. It is possible to set up a conversation in such a way as to disable the transformation. Thus in the Trollope quote above Augustus may be known both as a committed moral thinker and a sincere friend, and both these features may be sufficiently manifest earlier on in the conversation. Let's also assume for simplicity that he is widely known not to be guilty of any  $\phi$ -ing. Might there be, in these circumstances, a friendly and warm exchange of ideas about the morality of Felix's marriage plans? I can't see why not. But this tells in favour of the agonistic model. Augustus will have to make a serious effort to present his criticism as a friendly act. The very fact that a special effort should go into reconciling a moral criticism with friendly intents—this fact shows that the default context where a moral criticism occurs is an agonistic interaction. That is why in the actual case Felix instantly suspects hostility on Augustus' part, and that is why Augustus implicitly agrees with his interpretation and assures Felix that, far from moralising, he is speaking as a friend. As I will argue shortly, this idea of a default context is also key to solving Cohen's puzzle. And in Sect. 6 I'll argue that there is a deeper reason why moral criticisms normally belong in agonistic conversations.

Secondly, there are agonistic conversations where a riposte to a criticism would be out of place. Suppose that Gingrich has just emphasised his remorse for past indiscretions. If he now criticises Clinton for similar acts, the exchange likely continues to be agonistic. We don't necessarily infer that Gingrich has a purely moral concern in mak-

ing his criticism: he is probably still aiming to score rhetorical points. Still, Clinton can't very well accuse the opponent of hypocrisy with a simple riposte. The agonistic model can explain why. Since the riposte's purpose is to propose a counter-update of the common ground, it will violate the maxim of quantity. The information it is intended to convey is *already* in the common ground. On the other hand, it is possible that Gingrich's *remorse* itself was hypocritical. Though this case is not the focus of the present discussion (we are examining the pragmatics of criticism), the agonistic model can deal with it, too. Gingrich's presumed hypocrisy will be explained by citing his intent to use the statement of remorse as a move in an agonistic interaction. It is likely a defence against Clinton's future accusations, perhaps also an attempt to present himself to the audience as an individual concerned with morality.

You might object to the explanation just offered as follows. Suppose that Gingrich criticises Clinton for an indiscretion, although it is generally known that he himself is guilty of similar acts. Then, it seems, it would *not* be strange for Clinton to issue a riposte and accuse Gingrich of hypocrisy. Yet, just like above, the riposte appears to merely state the obvious and violate the maxim of quantity. Here I suggest to use some ideas of cognitive linguistics. In any minimally protracted conversation we can distinguish the semantic content currently in the field of the audience's attention, the content that was in that field earlier in the conversation, and then also the content merely in the common ground.<sup>24</sup> A new utterance brings its content into the focus of the audience's attention. This is so as long as the audience is prepared to consider at all what was just said. We also assume that the auditors have certain beliefs about Gingrich, and that these beliefs can motivate them to dismiss his utterance. But if they are merely part of the common ground, then they first have to be *activated* in the consciousness and become the focus of the auditors' attention. Cognitively, this carries activation costs. And on many occasions, the auditors would try to minimise the costs, for themselves and for *their* audience, and alleviate the load. As a result, entities referred to in Gingrich's utterance (e.g., components of Clinton's behaviour) would become the topic of the ongoing conversation which would attract further comments. It is then increasingly likely that the audience would form a negative view of Clinton *and* ignore the compromising information about Gingrich. The riposte is intended to prevent this outcome by activating the latter information in the consciousness of the auditors. The riposte, therefore, doesn't state the obvious. It rather makes the assumptions about Gingrich obvious *again*.

Before returning to Cohen's puzzle, let me elaborate the agonistic model a bit further. So far I have described agonistic conversations, like an imaginary one between Gingrich and Clinton, in terms of their common ground. Participants share the assumptions about the agonistic goals and intentions of the other participants and their own. But I also spoke freely of "transformations" and "fabrications" in the contexts where initially the participants didn't conduct an overtly agonistic conversation. Can this informal talk too be made more precise with the notions of the standard pragmatic account, like common ground updates and accommodation?<sup>25</sup>

<sup>24</sup> See Chafe (1994:53–56, 71–78), Langacker (2012:107–111).

<sup>25</sup> Standard pragmatic account: Stalnaker (2014).

Not quite. When, for example, you criticise me in the middle of an ostensibly friendly chat, I change many of my assumptions about what actually is going on in the conversation. Before, I thought we were friends sharing our problems and experiences with each other. Currently, I realise that you were trying to score rhetorical points against me, and that I was mistaken about the common ground all along. This change in my beliefs may *sometimes* be described as harmonising my presuppositions with yours and adding them to the common ground. I'm then "accommodating" the evidence, as I interpret it, of a defect in the context that your utterance has brought to light.<sup>26</sup> This is in a scenario where you too believe our exchange to be agonistic. But it is possible, indeed likely, that this is not what you *believe*. If pressed, you would deny that we are contesting some kind of asset. Unlike Augustus Staveley, you would stick to the moral idiom and still insist that you are speaking as a friend having no hostile intentions. Whether you *sincerely* believe that is a delicate ethical and psychological question. The least we can say is that many moral critics, including putative hypocrites, do not believe that they are party to an agonistic conversation (see also footnote 22).

Secondly, we can't use the standard account to explain why moral idioms should trigger such a radical overhaul of the common ground. Or rather, we can't use it beyond saying the obvious, that the overhaul is my rational response, as I see it, to the "manifest event" of moral criticism. Note too that triggering here is quite different from presupposition triggering: I'm reinterpreting the common ground not in order to make what you have said minimally intelligible, but to understand your broad motives and goals in the exchange.<sup>27</sup>

For a better explanation of the exchange between the critic and the criticism's target I propose to use the notion of frame.<sup>28</sup> This is a body of knowledge, produced by our familiarity with institutions and actual practice of language, that allows us to categorise interactions between speakers and hearers. Depending on the kind of interaction we expect a predictable back-and-forth made with similarly predictable lexical choices. On the other hand, particular lexical choices by the speaker confirm, challenge, or create anew the assumptions about the kind of conversation that is going on. Along with other cues, like gestures, intonations, facial expressions, they give us reliable if defeasible evidence about the goals and interests of the interlocutors. Our subsequent utterances are made in accordance with this understanding. Thus we may speak of "default frames" where particular lexical choices and other pragmatic cues normally, if not necessarily, belong.

We can now explain why a riposte is a felicitous reaction to moral criticism. When the conversation is already framed agonistically, its participants are expected to produce evidence and reasons to damage the interests of some other participants, generally of some of the people who are not party to the conversation. Moral criticisms play exactly this role. Your critic gives you "hard time" by damaging your prospects to secure some asset, likely by damaging your reputation.

<sup>26</sup> Accommodation: Stalnaker (2014:6, 47ff).

<sup>27</sup> See the instructive remarks in Stalnaker (2014:75ff).

<sup>28</sup> Frames and language: Fillmore (1975), Fillmore (1985). I ignore Fillmore's distinction between interactional and cognitive frames, and I don't endorse his stronger claims about semantic meaning being determined by frames. Other penetrating discussions: Goffman (1974:499ff), Gumperz (1982:130ff).

Similarly, a moral criticism in the middle of an ostensibly friendly interaction triggers our agonistic re-framing of the interaction in virtue of the lexical choices (as well as prosody choices in oral delivery). Expressions such as “You were morally wrong to  $\phi$ ”, “It was your duty to abstain from  $\phi$ -ing” make me reconsider our interaction. Until now I thought you had my interests at heart. After your criticism I see that you are aiming to damage them. My response is to answer in kind.

We are now in the position to solve Cohen’s puzzle (Sect. 3):

- (6) You were wrong to flirt, but I’m not in a position to criticise you.

Compare first these utterances:

- (13) a. You’ve put on a lot of weight. Just kidding!  
 b. I love hugging trees. No, seriously!

Informally, the speakers insert remarks (call them “metapragmatic comments”) to alter the “impression” created by the earlier utterance. They seem to say, “You may have the impression I was serious/joking, but I’m rather joking/serious.” Not that, with their earlier utterance, they implicated or presupposed any seriousness or joking. Instead, the earlier utterance, as normally used (i.e. in its default frame), would be a cue for the audience to understand that the speaker was serious (in 13a) or joking (in 13b). Metapragmatic comments correct this impression. The earlier utterance led the audience to believe that the present frame was serious/joking. The metapragmatic comment explicitly specified that the frame was not a default one, and that it was in fact joking/serious.

Consider now these utterances:

- (14) a. A: You were wrong to flirt. Speaking as a friend!  
 b. B: No offence, but you were wrong to flirt.  
 c. C: I’m not judging you, but you were wrong to flirt.  
 d. D: You were wrong to flirt. Just sayin’!

The remark in (14a) is readily intelligible if the default frame of “You were (morally) wrong to flirt” is one of animosity and conflict. To reframe this utterance a special effort is necessary—here, an explicit comment to indicate the friendly intentions of the speaker. In (14b) there is no explicit metapragmatic comment. Instead, the added qualifier serves the same purpose of reframing. *B* indicates his non-hostility, something that can’t be taken for granted absent the qualifier.

The most interesting cases for us are (14c)–(14d). If *C* is not judging, what *is* he doing? And if *D* is “just sayin’”, what’s he *not* doing? Following Cohen, we could take these locutions as evidence for the two distinct illocutionary acts of assertion and criticism to be performed with the same utterance “You were wrong to flirt.” The speaker in advance surrenders, for whatever reason, the authority of a judge and clarifies that he is only asserting. If, as argued earlier, this explanation doesn’t work, what’s an alternative?

Note that in its semi-judicial sense “judging” occurs in a situation of conflict between the “judge” and the “accused”, who is also a “defendant”. Judges can inflict punishments that damage the interests of the defendant—who “defends” himself before the judge, i.e. defends his interests. As in (14a), the accused is expected to understand the main utterance as made by someone with hostile intentions. Thus it’s not that *C* isn’t judging or criticising the subject’s flirting. It’s rather that his intention, as he clarifies, is not to use that criticism to punish the subject (where punishment is taken very broadly to include any purposeful damage of one’s interests). So the qualifier “not judging” is less specific than “speaking as a friend”, since *C* doesn’t claim to take the subject’s interests to heart. It is probably comparable to “no offence but”. “Just sayin’”, taken literally anyway, is similarly unspecific. *D* clarifies that he has no hostile intentions. To my ear, it also sounds as though *D* is not quite committed to the truth of his utterance. A possible follow-up could be an attribution of this assertion to someone else, like in the earlier example (7).

Cohen’s example (6) should be understood analogously. The speaker adds the qualifier as a cue to the audience that his utterance is not part of an agonistic exchange, and that he doesn’t pursue some interest at the expense of the person criticised. If, e.g., his own past flirting is known, he reasons that the auditors would frame his criticism as part of an agonistic exchange. The next move is likely to be a riposte. Either because his intentions are friendly from the start, or because he wishes to disable the riposte, the speaker seeks to re-frame his criticism. In short, he aims to inform the audience that the criticism does not occur in its default agonistic frame.

## 6 Moral criticism as side-taking

For all the benefits of the frame-theoretic approach, a vital explanatory link is still missing. Why, exactly, would the idiom of moral criticism belong in the agonistic frame? We seem to be painting an excessively cynical picture of moral condemnation. Many, I am sure, will be up in the arms.

True, they might say, on occasions moral criticisms and ripostes to them are deployed agonistically.<sup>29</sup> Political contexts are a prime example. But on many more occasions this won’t be so. We haven’t seen *why* we must embrace the second horn of Dworkin’s dilemma (Sect. 4). A few cherry-picked examples are no substitute for a general argument. The constructive conversations imagined by Dover remain a distinct *possibility*.<sup>30</sup>

To say anything conclusive on this matter it would be necessary to delve into the ethics of love and care and the nature of morals. I won’t do it here. Instead, I’ll finish with something far more modest and suggest a systematic empirical reason for linking moral criticisms to agonistic frames.

For starters, a few sociological observations:

<sup>29</sup> See, e.g., the concessions in Dover (2019a:41).

<sup>30</sup> Note that the constructive model also lacks an explanation why moral criticisms can regularly appear in “friendly” frames—why, in other words, they should *not* trigger agonistic transformation. Dover acknowledges this deficiency in (2019a:43), but offers no positive proposal.

[M]uch of talk takes place in the visual and aural range of persons who are not ratified participants and whose access to the encounter, however minimal, is itself perceivable by the official participants. These adventitious participants are “bystanders”. Their presence should be considered the rule, not the exception. The relation(s) among speaker, addressed recipient, and unaddressed recipient(s) are complicated, significant, and not much explored. An ideal in friendly conversation is that no one participant serve more frequently, or for a longer summation of time, in any one of these three roles, than does any other participant. In practice, such an arrangement is hardly to be found, and every possible variation is met with. (Goffman 1981:132–133)

Conversations are held often, even predominantly, in the presence and for the benefit of the third party (the “unaddressed bystanders”).<sup>31</sup> One-on-one friendly dialogues are rare and unstable. They would typically evolve either to include the bystanders, or become sermons, with one party usurping the speaker’s role. But why would a multi-party conversational setup be pervasive? A possible answer is that we are in a situation of conflict where the task is to recruit allies to our cause, with rhetoric a primary means of recruitment. Note too that the same purpose may be served in a dyadic conversation where the speaker is looking for sympathy from his sole interlocutor in the conflict with an absent agent (Goffman 1974:503). The next move is to speculate that moral criticisms are especially useful for these coalition-building efforts.<sup>32</sup> This speculation may be substantiated by the evolutionary argument in DeScioli and Kurzban (2009, 2013).

Let’s think of moral criticism—or “moral condemnation”, in DeScioli and Kurzban’s terminology—as an *evolutionary adaptation* designed to assist us in situations of conflict. Crucially, we count there not only combatants, but also observers (Goffman’s unaddressed bystanders). The observers’ task is to choose sides. The observers may use various strategies to achieve this, such as siding with the most powerful individual, or with the kin. But these strategies are based on the combatant’s identity and carry the risks of emerging despotism and escalating arms race, respectively.

A strategy that avoids both problems is “dynamic coordination”. Under this strategy, alliances are not static: they shift based on the actor’s performance. Moral cognition is functionally adapted to deal with forming alliances in this way. Generally, moral exchanges are represented as involving “perpetrators”, “victims”, and “judges” (condemners).<sup>33</sup> Each of these roles presents its own cognitive problems, and the three-party interaction may be examined from three different perspectives (DeScioli and Kurzban speculate further that these perspectives are in fact served by different cognitive modules). Perpetrators must weigh the costs of their actions against the revenge of the victims. Normative violations must be undetected, so far as possible. Intentions must be presented outwardly to the possible judges as benign. In addition, perpetrators must be able to compute the chances of retaliation and train themselves not to engage in the actions that are likely to prompt such a retaliation (i.e. to develop

<sup>31</sup> Goffman’s distinctions are refined in Levinson (1988).

<sup>32</sup> The “natural setting” of moral judgement: Haidt (2001:820).

<sup>33</sup> Judges may also praise, but as DeScioli and Kurzban note, condemnation prevails.

internal inhibitions against such actions). Victims, on the other hand, must minimise the costs of the hostile actions. Since one major source of that aid is from the judges, they have to present themselves *as* victims, and so to influence the judgement of the judges and their punishment routine.

As for the judges, they must decide on the nature of the action, determine the motives and goals of the perpetrators, and allocate punishment—namely, moral condemnation. They have to tread carefully to avoid (excessive) retaliation from the perpetrators and balance it against the potential rewards from other parties. From the judges' perspective, the conflict is represented as a collection of actions arranged on a scale of "wrongs". The moral(ised) scale of wrongs is negotiated and debated (possibly in advance), as is the classification in real time of particular actions in accordance with this scale. So understood, moral condemnation by an actor  $x$  serves as a signal for other parties (perpetrator, victim, other observers) that indicates which side  $x$  has joined.

This account of moral criticisms straightforwardly explains the perception that they are hostile acts. That's because, chances are, we have encountered them earlier on in our lives mostly in the situations of conflict where they were used to damage the criticised person's interests. This is no accident, since this employment matches the evolutionary function of moral condemnation, and of moral judgement generally (DeScioli and Kurzban 2018:181).

By the same token, this view fits well the thesis of Agonistic transformation. When a moral criticism is issued out of the blue, it is rational for us to attribute agonistic intentions to the speaker. The same goes for the situation of an avowedly friendly exchange. Faced with a moral criticism, you have a good reason, if a defeasible one, to reinterpret what was going on so far. You thought it was a friendly heart-to-heart chat, but it turns out your interlocutor approached the exchange agonistically.

How to think of the charge of hypocrisy? On DeScioli and Kurzban's view, since the criticism of  $x$  is a signal to the audience about  $x$ 's moral performance, the audience can use it as a reason to form a similar judgement and side against  $x$  in the conflict with the putative victim. The charge of hypocrisy is designed to undermine the signal's reliability by undermining the sender's honesty, where honesty is parsed as impartiality, a universal standard of moral judgement (2013:11). That moral judgements are *supposed* to be impartial follows from the way morality is designed to resolve conflicts: sides are chosen on the basis of the actor's observable behaviour, rather than his identity and his antecedent relation to the critic.

At the same time, according to DeScioli and Kurzban, hypocrisy and hypocritical condemnation are pervasive. Though moral judgements are required to be impartial, other evolutionary pressures are directed toward partiality, a behaviour favouring oneself and one's own kin (Kurzban 2010:206–217). Possibly a related reason is that moral judgements are cognitively economical when formed intuitively, rather than by calculation. Intuitive judgements are formed efficiently by identifying one's "selfish" preferences, rather than the preferences determined impartially.

This view of hypocrisy is open to two objections. (i) If hypocrisy is pervasive, the bystanders should be aware of this. Then moral condemnation won't be effective, since, according to the present view, it would normally be understood to be unreliable (Sect. 2). Ripostes would be idle, since, again, the critic's hypocrisy would be a default assumption. (ii) Moral criticism is supposed to be made for the benefit of bystanders.

Then how would an accusation of hypocrisy undermine its value in their eyes? Suppose that Gingrich criticises Clinton for adultery. Clinton responds by pointing out Gingrich's own adultery. Well, Gingrich may be partial or otherwise flawed, but so what? "Our subject tonight is Clinton's character," Gingrich could retort. "Whatever my own faults, put them aside." Of course, one response here is that, since he has no right standing, Gingrich's criticism simply doesn't count. But this idea we've already rejected (Sect. 3).

Both objections share the premiss that ripostes are aimed at undermining the credibility of the critic and the criticism. Hence they may be dealt with swiftly by using the agonistic model where this premiss is dropped. In the Clinton-Gingrich scenario, the subject of the conversation isn't just Clinton's character. We aren't examining it in abstract, *sub specie aeternitatis*. We are in a situation of rivalry between Clinton and Gingrich, and to side against Clinton is to side with Gingrich. If you, as a bystander, think that adultery is a reason to side against Clinton, then you have exactly the same reason to side against Gingrich.

Ripostes, therefore, are aimed not at undermining the critic's credibility, but at creating an alternative way to coordinate alliances among the bystanders. DeScioli and Kurzban's view of hypocrisy doesn't sit well with their own theory of moral condemnation. An agonistic explanation is in fact more in line with the latter.

## 7 Review

To sum up, I have argued against several extant ideas about the pragmatic role of moral criticism and the corresponding views of hypocrisy. Moral criticism does not presuppose a moral commitment, nor does it conversationally implicate it. Its illocutionary force can't properly be described as depending on a special moral authority of the speaker. I argued further that it is best to locate moral criticism in a situation of rivalry over certain assets where moral considerations influence the respective chances of the rivals. Accusations of hypocrisy aim to disable the damaging effects of the criticism on the chances of the individual criticised.

**Acknowledgements** I'm grateful to an anonymous referee for useful comments on an earlier draft of this paper.

## References

- Alston, W. P. (2000). *Illocutionary Acts and Sentence Meaning*. Ithaca: Cornell University Press.
- Austin, J. L. (1975). *How to Do Things with Words*. Clarendon Press, Oxford, 2nd edn. Edited by J. O. Urmson and M. Sbisà.
- Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge: The MIT Press.
- Berkovski, Y. S. (2022). Mandeville on self-liking, morality, and hypocrisy. *Intellectual History Review*, 32(1), 157–178.
- Brown, P., & Levinson, S. C. (1987). *Politeness*. Cambridge: Cambridge University Press.
- Chafe, W. L. (1994). *Discourse, Consciousness, and Time*. Chicago: The University of Chicago Press.
- Cohen, G. A. (2013a). Casting the first stone. In *Cohen* (2013).

- Cohen, G. A. (2013c). Ways of silencing critics. In *Cohen* (2013).
- Cohen, G. A. (2013). *Finding Oneself in the Other*. Princeton: Princeton University Press.
- Crisp, R., & Cowton, C. (1994). Hypocrisy and moral seriousness. *American Philosophical Quarterly*, 31(4), 343–349.
- DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, 119, 281–299.
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139(2), 477–496.
- DeScioli, P., & Kurzban, R. (2018). Morality is for choosing sides. In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology*. New York: Guilford Press.
- Dougllass, R. (2022). Bernard Mandeville on the use and abuse of hypocrisy. *Political Studies*, 70(2), 465–482.
- Dover, D. (2019). Criticism as conversation. *Philosophical Perspectives*, 33, 26–61.
- Dover, D. (2019). The walk and the talk. *Philosophical Review*, 128(4), 387–422.
- Dworkin, G. (2000). Morally speaking. In E. Ullmann-Margalit (Ed.), *Reasoning Practically*. Oxford: Oxford University Press.
- Fillmore, C. J. (1975). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1), 20–32.
- Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2), 222–254.
- Fritz, K. G., & Miller, D. (2018). Hypocrisy and the standing to blame. *Pacific Philosophical Quarterly*, 99, 118–139.
- Goffman, E. (1974). *Frame Analysis*. New York: Harper & Row.
- Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Grice, H. P. (1989). Utterer's meaning and intentions. In *Studies in the Way of Words*. Harvard University Press.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66(3), 377–388.
- Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.
- Haidt, J. (2001). The emotional dog and its rational tail. *Psychological Review*, 108(4), 814–834.
- Isserow, J., & Klein, C. (2017). Hypocrisy and moral authority. *Journal of Ethics and Social Philosophy*, 12(2).
- Kurzban, R. (2010). *Why Everyone (Else) is a Hypocrite: Evolution and the Modular Mind*. Princeton: Princeton University Press.
- Langacker, R. W. (2012). Interactive cognition: Toward a unified account of structure, processing, and discourse. *International Journal of Cognitive Linguistics*, 3(2), 95–125.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (1988). Putting linguistics on a proper footing. In P. Drew & A. Wootton (Eds.), *Erving Goffman: Exploring the Interaction Order*. Polity Press Polity Press Polity Press Polity Press.
- Lippert-Rasmussen, K. (2020). Why the moral equality account of the hypocrite's lack of standing to blame fails. *Analysis*, 80(4), 666–674.
- Mandeville, B. (1924). *The Fable of the Bees*, volume 1. Clarendon Press. Edited by F. B. Kaye.
- Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. Oxford: Oxford University Press.
- Piovarchy, A. (2020). Situationism, subjunctive hypocrisy and standing to blame. *Inquiry*. Forthcoming.
- Portner, P. (2016). Imperatives. In M. Aloni & P. Dekker (Eds.), *The Cambridge Handbook of Formal Semantics* (pp. 593–628). Cambridge: Cambridge University Press.
- Potts, C. (2015). Presupposition and implicature. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory* (pp. 168–202). Hoboken: Wiley Blackwell.
- Riedener, S. (2019). The standing to blame, or Why moral disapproval is what it is. *Dialectica*, 73(1–2), 183–210.
- Rossi, B. (2018). The commitment account of hypocrisy. *Ethical Theory and Moral Practice*, 21, 553–567.
- Rossi, B. (2021). Hypocrisy is vicious, value-expressing inconsistency. *Journal of Ethics*, 25, 57–80.
- Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning*. Blame: Belknap Press, Cambridge, MA and London.
- Searle, J. R., & Vanderveken, D. (1985). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.
- Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford: Clarendon Press.
- Shklar, J. N. (1979). Let us not be hypocritical. *Daedalus*, 108(3), 1–25.
- Shoemaker, D., & Vargas, M. (2021). Moral torch fishing: A signaling theory of blame. *Noûs*, 55(3), 581–602.

- Simons, M. (2006). Foundational issues in presupposition. *Philosophy. Compass*, 1(4), 357–372.
- Stalnaker, R. C. (2014). *Context*. Oxford: Oxford University Press.
- Szabados, B. (1979). Hypocrisy. *Canadian Journal of Philosophy*, 9(2), 195–210.
- Todd, P. (2019). A unified account of the moral standing to blame. *Noûs*, 53(2), 347–374.
- Tonhauser, J., Beaver, D., Roberts, C., & Simons, M. (2013). Toward a taxonomy of projective content. *Language*, 89(1), 66–109.
- Trollope, A. (1985). *Orley Farm*. Oxford University Press. Edited by D. Skilton.
- von Fintel, K. (2004). Would you believe it? The King of France is back! In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and Beyond*. Oxford: Oxford University Press.
- Wallace, R. J. (2010). Hypocrisy, moral address, and the equal standing of persons. *Philosophy & Public Affairs*, 38(4), 307–341.
- Williamson, T. (2000). *Knowledge and Its Limits*. New York: Oxford University Press.
- Wolf, S. (2011). Blame, Italian style. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. Oxford: Oxford University Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.